PRIZE PUBLICATION FUND
VOL. XX

C-59

Turkish and Mongolian Studies

By

SIR GERARD CLAUSON

Клаусон Ж.
Тюрковедение и монголоведение. Лондон,
1962, 267с.

Published by

THE ROYAL ASIATIC SOCIETY OF
GREAT BRITAIN AND IRELAND

and sold by its agents

LUZAC AND COMPANY LIMITED
46 Great Russell Street, London, W.C.1

1962

CONTENTS

	page
Foreword	VII
Chapter I. The early history of the Turkish-speaking tribes	1
Chapter II. The evolution of the Turkish languages	33
Chapter III. The transcription of Turkish languages	55
Chapter IV. The arrangement of words in a historical dictionary of Turkish	62
Chapter V. The evidence regarding the phonetic structure of pre-eighth century Turkish ...	67
Chapter VI. The structure of the word in pre-eighth century Turkish	135
Chapter VII. The suffixes in pre-eighth century Turkish	140
Chapter VIII. The reconstruction of the phonetic structure of pre-eighth century Turkish ...	160
Chapter IX. The later history of the Uyğur alphabet	175
Chapter X. The phonetic structure of pre-thirteenth century Mongolian	188
Chapter XI. The relationship between Turkish and Mongolian	211
Epilogue	248
Addendum	256
Index	257

FOREWORD

Kimiŋ yaşı altmış tüketse sakış
Tatığ bardı andın yayı boldı kış

When a man has reached the age of sixty, the savour of life has gone, and his summer has become winter.

Kutadğu: Bilig, verse 367.

The days of our age are threescore years and ten; and though men be so strong that they come to fourscore years; yet is their strength then but labour and sorrow, so soon passeth it away and we are gone.

Psalm 90, verse 10 (Prayer Book version).

There is a gracious practice in the learned world, whereby the friends and pupils of a distinguished scholar celebrate his seventieth birthday by clubbing together and producing a volume of articles dedicated to him and called alternatively Charisteria, Symbolae, Festschrift, Armağan or, in English, more prosaically Anniversary Volume.

It would be uncharitable, and indeed unjust, to suggest that one of the motives at the back of the contributors' minds is a vague hope that the old gentleman will take this as a discreet hint that he is past work and ought to retire and make way for others. Nevertheless the fact remains that a seventieth birthday is something in the nature of a milestone indicating that from then onwards the road can only go downhill. Not being myself an established denizen in the learned world—at the best I might be described, in the language of the National Insurance Acts, as a late entrant, at the worst as a mere gate-crasher—it would be wildly inappropriate for me to be given the compliment of an Anniversary Volume; and so, being immune from professional competition, I felt that I might suitably reverse the process and celebrate my seventieth birthday, a little in arrear, by presenting to my many friends—I never had any pupils—a report on some of the conclusions which I have reached as a result of nearly ten years' intensive study of the early history of the Turkish language. There is less vanity and more good sense in this than might at first sight appear. We are all too familiar

with the phenomenon of scholars who absorb knowledge like a sponge all their lives and in the end take it with them to the grave, leaving little or nothing to show for it. My purpose in presenting what is, I hope, only an interim report is to put together and present in logical order some of the facts about the Turkish and Mongolian languages which I have learnt during these past ten years, and to indicate certain directions in which I suggest that further research can profitably be directed either by myself or by others.

Since writing the first draft of this book I have given some thought, but perhaps not enough, to my possible audience, and endeavoured by judicious revision to make it serve three rather different purposes. I hope that in doing so I have not fallen between three stools.

There is a real need for an introduction to Turkish studies, suitable for elementary students of any of the Turkish languages, which will give them some idea of the evolution of the Turkish language group as a whole, help them to see what position the language they are proposing to study, perhaps Republican Turkish or one of the Turkish languages spoken in the Soviet Union, occupies in that group, and encourage them to broaden and deepen their interest in the group as a whole. One of my objects has been to meet this need to a modest degree.

There is also a real need to re-examine the transcriptions of the early Turkish texts, down to, say, the eleventh century inclusive, to correct them in certain respects, and so to establish the phonetic structure of the languages in which they were written, and from these to deduce the phonetic structure of the earliest form of Turkish which we can visualize. This need too I have endeavoured to meet, but the subject is a difficult and technical one and I am afraid that the elementary student, if he gets that far in the book, will find it very heavy going.

Finally it would be nice to bring the question of the "Altaic theory," which has now bedevilled Turkish studies for many years, to a head, and so, if possible, to settle once

and for all whether the Turkish and Mongolian languages are genetically related to one another; in Chapter XI I have stated the reasons for which I believe that they are not. This and the two preceding chapters are designed for yet another audience, perhaps not so much the Turcologists as the Mongolists, who will find little of interest for their own discipline in the rest of the book.

It may perhaps be of interest to explain how I came to write a book combining these three rather disparate themes. I first acquired an affection for the Turkish language at the age of fifteen, but the pressure of official and other duties prevented me from devoting much time to its study for the next forty-five years. However the old allegiance never wavered, and when I retired from public service at the end of 1951, I resolved to devote my declining years to the history of the Turkish language. In fact, since that date the whole of my time that has not been occupied by domestic, social or business matters has been devoted to that study. The result of this work has been a trickle of occasional papers which are listed at the end of this Foreword for convenience of reference and a larger volume of work which is not yet in a form suitable for publication. The advantages of pursuing a line of study without a salary and consequently without a master are incalculably great. Undistracted by the demands of pupils and the exigencies of a programme laid down from above, I have been able to follow my studies in whatever direction they led me, starting and discontinuing particular lines of research as circumstances seemed to demand.

When I retired at the end of 1951 my first resolve was to compile a new and better "Radloff," a historical dictionary of all the Turkish languages from the earliest times to the present day, excluding loan words from other languages. This project petered out in the sands at the end of about six years. It became quite clear that it was impracticable for me, and indeed probably impracticable for any single scholar, even if he started in his youth and not, as I did, at the age of sixty, to compile such a work

single handed, and I began increasingly to doubt whether it would be of real utility even if it were compiled. My reasons are stated in Chapter II of this book. Even as I wrote, new authorities came pouring in, so that each article became more and more incomplete directly it was written. By that time I had finished, as it had already become clear quite imperfectly, the work on words beginning with vowels, and was nearly half way through the work on words beginning with **ç-** and **c-**, a group which I had chosen because I thought, in my innocence, that it might be a fairly easy one after the matted jungle of words beginning with vowels. And so I started all over again on a much more modest scale to compile a historical dictionary of the early Turkish languages including, with a few exceptions, only words which are known to have existed before the Mongolian expansion in the early thirteenth century. This work is still proceeding and I reckon that over half the first draft is now done.

This does not mean that I regard the labours of those first six years as wasted; quite the contrary. It was an invaluable experience to take word after word, to establish its first appearance and meaning, to trace its history forward from that point and find out what later forms it assumed and what finally became of it. This is in fact the only way in which to get a real feel for the history of the language. It taught me at least four important lessons.

The first in chronological order was that Turkish dictionaries contain many supposedly Turkish words which never really existed at all. My conclusions on this subject were embodied in *Turkish ghost words*. This paper contains some minor inaccuracies, but I see nothing of substance to alter in it, although, as I shall mention below, I have added some elaborations in another place.

A by-product of this discovery was the realization that there was a large concentration of these "ghost words" in the dictionaries of Çağatay, owing chiefly to the fact that scholars who worked in this field had never had the advantage of basing their studies on the only reliable

dictionary of Çağatay compiled by an eastern scholar, the *Sanglax* of Mirza Muḥammad Mahdī Xān. My early studies of this manuscript were pursued with the help of manuscript Or. 2,892 in the British Museum, which is a sort of calligraphic nightmare, but I was so fortunate as to find an incomparably better manuscript in the possession of the E. J. W. Gibb Memorial Trustees, and even more fortunate in being able to persuade them to publish a facsimile of it with an introduction and indices by myself. This meant taking a good deal of time off from my regular lexicographic work, but I do not regret it. The publication of this book has, I hope, laid a more solid foundation for Çağatay studies in future. I hope too that the further elaboration of my remarks on the genesis of ghost words and the explanation of the way in which the compilers of Çağatay dictionaries worked, which I included in my introduction, will stimulate other scholars in the pursuit and elimination of more ghost words.

The second lesson which I learnt, more particularly from the *Sanglax*, was that there is a much greater proportion of Mongolian loan words in most mediaeval and modern Turkish languages, but more particularly those in the north-eastern group (the languages of southern Siberia, including Tuva), the north central group (Kazax and Kırğız) and in Çağatay than had hitherto been suspected. It has of course long been realized that all the languages spoken by Moslem Turks are full of Arabic loan words, and some of them, more particularly in their literary forms, full of Persian loan words as well, but the existence of this massive volume of Mongolian intruders seems somehow to have escaped notice, or, if noticed, to have been regarded by those who accept the Altaic theory as evidence of a common " Altaic " heritage in both language groups. So far as Çağatay is concerned, I devoted a good deal of attention to the subject in the introduction and relevant indices to the *Sanglax*.

The third lesson which I learnt was that the phonetic structure of the earliest kind of Turkish that we can

reconstruct, that spoken appreciably earlier than the eighth century, the date of the earliest substantial remains of Turkish which have come down to us, was a good deal more elaborate than had hitherto been realized, both as regards the general repertoire of sounds, and also more particularly as regards the sounds which could be used at the beginning of a word. This first came actively to my attention when I was working on words beginning with **ç-** and **c-**, and I embodied some preliminary observations on the subject in *The Turkish Y and related sounds*. I returned to this subject recently in *The initial labial sounds in the Turkish languages*. Another by-product of my lexicographic labour at this period was *The Turkish numerals*, which *inter alia* established some unexpected facts about the phonetic structure of some numerals, and showed that one class of collective numerals was a Turco-Mongolian hybrid and not pure Turkish.

The central point of the Studies which follow is Chapter VIII, in which I have embodied what I have so far learnt about the phonetic structure of pre-eighth century Turkish. It is not intended to replace any existing work; for the most part it deals with an earlier stage of the language than any hitherto studied, but it is intended to provide a solid basis for the re-examination, and where necessary revision, of the conclusions set out in such classical works on the phonetic structure of the modern Turkish languages as:—

W. Radloff, *Phonetik der nördlichen Türksprachen*, Leipzig, 1882;

N. F. Katanov, *Opyt izsledovaniya Uryankhayskago Yazyka*, Kazan, 1903;

M. Räsänen, *Materialen zur Lautgeschichte der Türkischen Sprachen*, Helsinki, 1949;

L. Bazin, *Structure et tendences communes des langues turques* (*Sprachbau*) in *Philologiae Turcicae Fundamenta*, Wiesbaden, 1959;

N. A. Baskakov, *Tyurkskie Yazyki*, Moscow, 1960.

The fourth lesson which I learnt was that the early Turkish vocabulary had an extraordinarily tenacious hold

on life. There are a good many words which occur in the earliest languages, especially Uyğur and Xākānī, but do not seem to be noted as occurring in any later literary text or in any dictionary of the literary languages, and yet have survived in the spoken languages and so can still be found in V. V. Radloff *Opyt slovarya tyurkskikh narechiy*, St. Petersburg, 1888-1911 (cited as *Opyt*), as existing in languages of the north-eastern group, or in the *Türkiyede halk ağzından söz derleme dergisi*, Istanbul, 1939 ff. (cited as SDD) as existing in the contemporary spoken language of the Turkish Republic. It is very probable that further search in other modern Turkish languages would prove that other early words which are now supposed to be obsolete have in fact survived somewhere.

As a young man I had always accepted the theory that the Turkish and Mongolian languages were genetically related. It seemed *prima facie* probable, but I was not greatly moved by the subject; it was Turkish, and not Mongolian, that interested me. But I did accept it, and so when a Romanized text of the *Secret History of the Mongols*, a work that I did expect to interest me, became available, I tried to read it. I did not begin to understand it, and I could find nothing Turkish about the language in which it was written. And so I came to the conclusion that the theory that the Turkish and Mongolian languages were genetically related—the Altaic theory—was almost certainly wrong and lost all interest in the subject, except that later it strengthened my resolve to sort out the Mongolian loan words in Turkish and exclude them from my dictionary as a foreign element. And there the matter would have rested if I had not providentially been aroused by a rather brash attack on my old friend Prof. K. Grønbech in Central Asiatic Journal II 1, by a young reviewer of his and J. R. Kruger's excellent little hand-book *An introduction to Classical (Written) Mongolian*, Wiesbaden, 1955. The Professor, he said, should not be so ill-advised as to put into the heads of young students such a dangerous and incorrect idea as one that the Turkish and Mongolian languages

might not be genetically related, or, if he did, should give a reasoned statement of his views on the subject. By this time our old friend, though we did not know it at the time, had already fallen a victim to the deadly disease which so prematurely deprived us of the pleasure of his company, and in any case I felt sure that as a professional Orientalist he was much too busy to devote valuable time to refuting ill-informed critics. And so I thought, more particularly since a recent study of Tuvan had convinced me of the enormous volume of Mongolian loan words in some Turkish languages, that being my own master I could usefully change course for a time and subject the Altaic theory to critical analysis. My preliminary conclusions were embodied in *The case against the Altaic theory*. By the general conclusions in that paper I am still ready to stand firm, but it undoubtedly contains errors of detail. I have since dealt with various aspects of the subject in three further papers and my final conclusions will be found in Chapter XI of this book.

Study of some aspects of this question convinced me of the necessity for a new examination of the early history of the Turkish-, Mongolian- and Tungus-speaking tribes, a subject which I was ill-qualified to examine. However, with the help of my friend Prof. E. G. Pulleyblank I produced *Turk, Mongol, Tungus* and I have returned to this subject in Chapter I of this book. A great deal more work remains to be done on this subject, but it will have to be done by someone with an intimate knowledge of the Chinese authorities and the ability to discover and utilize texts which are not yet available in translation.

Another by-product of my Mongolian studies was *The hP'ags-pa alphabet*. This was in a sense a return to an old love; as long ago as 1929 S. Yoshitake and I had written a joint paper on the phonetic values of two letters of this alphabet.

This in its turn had been the by-product of a joint study on which the late Prof. F. W. Thomas and I had been engaged of some Chinese texts in Tibetan script. The

matter interested me because I hoped to get from these phonetic transcriptions of eighth century Chinese some light on the early Turkish names and titles which appear in Chinese transcription in contemporary Chinese records. That this study produced some useful results will appear in Chapter V. It was primarily to establish the exact pronunciation of Turkish loan words in thirteenth and fourteenth century Mongolian that I returned to the study of the hP'ags-pa alphabet.

This report, as I have already said, is I hope only an interim one. I still regard early Turkish lexicography as my principal task, but as it proceeds I hope that it may throw up material for further occasional papers.

I cannot conclude this Foreword without expressing my thanks to my many friends who have contributed ideas which have gone to the making of this book. They include scholars not only in this country but also in half-a-dozen foreign countries, and more particularly Hungary and the Soviet Union, with whom I have discussed individual problems, orally at Oriental Congresses and by correspondence between Congresses. Of scholars in this country I must particularly thank Edwin Pulleyblank, the Professor of Chinese at Cambridge University, whose help in connection with Chapter I has been invaluable; Victor Menage, Lecturer in Turkish at the School of Oriental and African Studies, University of London, who read the whole book and gave me the benefit of his help both as a meticulous copy reader, and as a scholar with a wide knowledge of Turkish and other languages; and Charles Bawden, Lecturer in Mongolian at the same School, who closely examined the last three chapters. By their combined efforts I have been saved from a number of obvious errors, and others not so obvious. That other errors still remain in spite of their help I have very little doubt. All that I can do is to repeat the poignant words (in 2 v.6 of the facsimile) of the author of the *Sanglax*:—*hargāh quṣūrī wa futūrī wāqi' ṣuda bāṣad, tawaqqu' az nāẓirān ān ki ba-xāma-i luṭf iṣlāḥ numāyand*, " wherever there are errors or defects, I beg

those who observe them to correct them with the pen of kindness."

LIST OF PUBLICATIONS

F. W. Thomas and G. L. M. Clauson. *A Chinese Buddhist text in Tibetan writing*, Journal of the Royal Asiatic Society, July 1926.

ditto. *A second Chinese Buddhist text in Tibetan writing*, J.R.A.S., April 1927.

A hitherto unknown Turkish manuscript in "Uighur" characters, J.R.A.S., January 1928.

S. Yoshitake and G. L. M. Clauson. *On the phonetic value of the Tibetan characters ['] and [h̲] and the equivalent characters in the h̲P'ags-pa alphabet*, J.R.A.S., October 1929.

The geographical names in the Staël-Holstein scroll, J.R.A.S., April 1931.

Communal Land Tenure, Food and Agricultural Organization of the United Nations, 1953.

Turkish ghost words, J.R.A.S., October 1955.

A note on Qapqan, J.R.A.S., April 1956.

The case against the Altaic theory, Central Asiatic Journal II 3, 1956.

À propos du manuscrit Pelliot Tibétain 1283, Journal Asiatique, 1957.

The Turkish Y and related sounds, Studia Altaica, Ural-Altaische Bibliothek, 1957.

The Ongin inscription, J.R.A.S., October 1957.

Araba, Encyclopedia of Islam, Vol. I, 1957.

Türkçe sözler üzerinde araştırmalar, Jean Deny Armağanı, Ankara, 1957.

The Turkish numerals, J.R.A.S., April 1959.

The h̲P'ags-pa alphabet, Bulletin of the School of Oriental and African Studies XXII, 2, 1959.

The earliest Turkish loan words in Mongolian, C.A.J. IV 3, 1959.

The Turkish elements in 14th century Mongolian, C.A.J. V 4, 1960.

[illegible]lān, a Persian guide to the Turkish language, by Muḥammad Mahdī Xān. Facsimile text, with an introduction and indices by Sir Gerard Clauson, E. J. W. Gibb Memorial, New Series XX, 1960.

[illegible], Mongol, Tungus, Asia Major VIII, 1961.

[illegible] Beshim-Suyab, J.R.A.S., April 1961.

[illegible] initial labial sounds in the Turkish languages, B.S.O.A.S. XXIV 2, 1961.

[illegible]tes on the "Irk Bitig", Ural-Altaische Jahrbücher XXXIII 3-4, 1961.

CHAPTER I

THE EARLY HISTORY OF THE TURKISH-SPEAKING PEOPLES

Al-turk fi'l-aṣl 'iṣrūna qabīla . . . wa kull minhā buṭūn lā yuḥṣīhim illā'llāh.

"The Turks were originally twenty tribes . . . and each of these tribes God alone can count how many clans."

Kāşğarī, I 28.

An essential preliminary to any study of the history of the Turkish language is a study of the history of the peoples who are known or believed to have spoken that language. Most of the readers of this book will no doubt already be pretty familiar with the subjects with which it deals, but some of them may be unprejudiced by any previous knowledge of the subject, and for their benefit it might be useful to set out here succinctly and in simple language the processes by which it can be proved, or at any rate conjectured with some confidence, that particular peoples whose names will occur frequently in this book spoke a Turkish language.

As will appear later, we have inscriptions or documents in some form of Turkish which can be firmly, or almost firmly, dated to every century from the present day back to the eighth century A.D.? Some of the documents may perhaps be copies of texts composed a little earlier, but probably not earlier than the seventh century. That is the extent of our direct knowledge. To reconstruct the history of the Turkish language and the peoples who spoke it before that date we must have recourse to indirect evidence. The process is one of pushing back from the known to the unknown with the help of this indirect evidence hoping that what will emerge from it will be a convincing, or at any rate a plausible, picture of this earlier history. At the same time it is important not to underestimate the difficulties even on the Turkish side. During the thirteen centuries for which we have direct knowledge there has been a slow but steady change in the sounds existing in the Turkish

languages and in their grammars and vocabularies. It would be unreasonable to suppose that similar changes did not take place in the preceding centuries. Indeed we can prove in one way or another that certain sounds and words existed in pre-eighth century Turkish which had by that time become obsolete; no doubt there were others of which we have not got even indirect evidence.

Our indirect evidence can be summarized under the following heads:—

(1) peoples whose languages survive in eighth and ninth century material are mentioned in history under the same names two or even more centuries earlier;

(2) historical records in other languages, mainly Chinese and (Byzantine) Greek, which are earlier than the eighth century, or at any rate based on lost documents earlier than that date, contain the names of peoples who were, we know, speaking Turkish at a time for which we have direct evidence, and this is sometimes confirmed by the quotation of words used by those peoples at that remote period, which are demonstrably Turkish and sometimes in forms earlier than those of the eighth century;

(3) we have similar demonstrably Turkish vocabulary material from this earlier period attributed to other named peoples, whose names are familiar but of whose languages no continuous texts survive;

(4) we have in the Chinese records statements to the effect that certain peoples spoke the same language as other peoples who are known to have spoken Turkish, and a different language from that of other peoples who are known not to have spoken Turkish, and also that peoples of whose languages specimens have survived were "descended" from certain earlier peoples, and so may be presumed to have spoken the same kind of language as those peoples.

The Byzantine records give us some invaluable information about certain Turkish tribes at a very early period, but this information is scanty. Substantially we are dependant

on the Chinese records for reconstructing the history of the Turkish peoples and their language prior to the eighth century. At the same time it would be ingenuous to suppose that these Chinese records make easy reading or can be easily interpreted. It is, in fact, extremely difficult for anyone who cannot read them in the original, and not much easier for anyone who can, to be sure that he has interpreted them correctly, even when he has good translations at his disposal. This arises not only from the fact that these texts are unusually difficult to understand, so that there are often important differences between two translations, both by competent scholars, of the same Chinese text, but also from the fact that the relationship between different texts concerned with the same subject matter, and more particularly the various histories covering the period from the beginning of the third century A.D. onwards, is unbelievably complicated, particularly so far as accounts of peoples and events beyond the Chinese frontier are concerned. A vivid account of the difficulties which have to be faced will be found in the Textual Appendix to R. A. Miller's *Accounts of western nations in the history of the Northern Chou dynasty*, University of California Press, Berkeley and Los Angeles, 1959. Most of the translations which we have for this period are translations of chapters from individual Dynastic Histories like the *Wei Shu* and the *P'ei Shih*, but, as Miller points out, in nearly every case such a text is derived from some earlier authority—for example several chapters of the *Wei Shu* were lost at a very early date and later replaced by relevant extracts from the *P'ei Shih*—or from some lost text which may have survived in a more satisfactory form in some other book, perhaps not even a Dynastic History. A great deal more work will have to be done in the way of critical editions with translations before we can be sure that we have before us all that the Chinese wrote about Turkish peoples in its earliest and most authentic form. Even when these texts have been assembled we shall still have to take the foibles of those who wrote them into account, the use

of standard phrases in describing almost any kind of "barbarian" and the fact to which W. Eberhard drew attention in Chapter 25 of *Das Toba-reich nord Chinas*, Leiden, 1949, that the Chinese historiographer's purpose was often not so much to give a dispassionate account of past events as to give the account a slant which pointed towards what would, in his opinion, be the most sensible direction of future policy. Thus some of the most vivid stories about the "barbarians," for example that about Mao-tun's rise to power, may be no more than exercises in fiction designed to point a moral.

There is another factor to be taken into account when Chinese scriptions of foreign names and words are used as philological evidence. The Chinese language, particularly so far as its phonetics are concerned, was a good deal less static even than Turkish. The Chinese themselves were very conscious of this fact, and later commentaries on the early records contain numerous statements to the effect that some Chinese character in the scription of a foreign name or word had a sound different from its normal one; but as they were working entirely with their own script, that is without any kind of phonetic alphabet, notes of this kind seldom carry us very far. For a good many years now Sinologists from other countries, and indeed some Chinese scholars familiar with alphabetic scripts, have been endeavouring to reconstruct the phonetic structure of earlier forms of Chinese, specifically dialects spoken in particular areas at particular dates. The most distinguished worker in this field has been Prof. B. Karlgren, who in a series of masterly works, culminating in *Grammata Serica*, Stockholm, 1940, revised as *Grammata Serica Recensa*, Stockholm, 1957, has produced a reconstruction of the phonetic structure of two early dialects, that called by him "Ancient Chinese," the language spoken in Chang-an, the capital of China, in the late sixth and early seventh century A.D., and that call by him "Archaic Chinese," the language spoken in the Chou capital nearly fifteen hundred years earlier. It has always been accepted that neither of these

dialects was spoken all over China at the dates stated. For example the transcriptions of Chinese in Tibetan script which Prof. F. W. Thomas and I published over thirty years ago and on which Prof. Walter Simon has since published a series of important monographs, although they are roughly contemporary with "Ancient Chinese," represent a dialect with some important phonetic differences. For example the sound which is represented by final *-t* in Ancient Chinese is habitually represented by *-r* in these texts and initial *n-* and *m-* by *nd-* and *mb-*. But the most important deficiency in Prof. Karlgren's work, impressive as it is, is the enormous gap in time between his two dialects. Since then other scholars have been striving to find means of filling this gap to some extent, and in particular Prof. Pulleyblank has been working on the problem of Chinese scriptions of foreign names and words during the Han period, that is roughly the four centuries centring round the turn of the Christian era. It would be quite improper for me to anticipate his conclusions, which will, I hope, appear in print before this book, but he has been so good as to discuss with me at length a number of problems of transcription and historical interpretation which have confronted me, and it will not, I think, be improper for me to summarize the lessons which I have learnt from these valuable discussions.

The first was that in the first millennium A.D. there were, as there still are, several dialects of Chinese with more or less widely divergent phonetic structures, so that a Southern Chinese writer would use different characters from a more or less contemporary Northern Chinese writer to represent the same foreign name or word, and both writers might use the same character but to represent different foreign sounds. The second was that the phonetic structure of all these dialects was altering fast between the beginning of the Han period in 202 B.C. and the T'ang period beginning in A.D. 620, so that even in the same area different characters would have been used to represent a foreign word in the second century B.C. from those used to

represent the same word eight or nine centuries later; indeed during this period the representation might have changed more than once. It will readily be understood that since a parallel alteration in the phonetic structure of Turkish itself was no doubt taking place during this period, it is supremely difficult to be sure that a Chinese scription of the early Han period represents a Turkish word of which the earliest form actually known to us dates to no earlier than the eighth century A.D. The third was that foreign tribal names which the Chinese used cannot be confidently assumed to have had the same ethnic connotation at different periods, and this for two reasons. The first is that that this is very often certainly not the case. It is notorious that the classical authorities, Greek and Roman, used the name " Scyth " for a number of " barbarian " tribes which had no right to that name—some of them were not even Iranian—and that the Byzantine authorities used the name " Turk " in the same indefinite way, at one period specifically for the Magyars. It is less notorious, but equally true, that at different times the Chinese used words like *hu*, " barbarian," and tribal names like Hsiung-nu for peoples of different ethnic origins. The second reason is that in conditions of steppe life the ethnic constitution of a particular horde or confederation might alter quite considerably as time went on, even though the name remained unchanged. Finally Prof. Pulleyblank pointed out to me that the Chinese records are not a beautifully coordinated and self-consistent whole. A later writer might well not be familiar with earlier works, though very often he was and used them without acknowledgement in a very uncritical way. Thus different authorities often give quite different accounts of the " origin " of a particular tribe. Sometimes two authorities can be coordinated and shown to tell the same story but in different words; but sometimes the two are quite irreconcilable, and one, or perhaps both, must be false. With this discouraging introduction I will now do my best to probe into the obscurities of Turkish prehistory.

The principal authorities containing translations of relevant Chinese texts which I have used in compiling this summary are:—

N. Ya. Bichurin (Père Hyacinthe or Yakinf), *Sobranie svedeniy o narodakh obitavshikh v Sredney Azii v drevniya vremena*, St. Petersburg, 1851, republished by the Soviet Academy of Sciences, Moscow, 1950 (cited as Bichurin *op. cit.*), a remarkable production for its period covering much more ground than any other single authority, and still containing some material not otherwise available, but with serious defects, some of which have been corrected in N. V. Kyuner, *Kitayskie izvestiya o narodakh yuzhnoy Sibiri, tzentralnoy Azii i Dal'nego Vostoka*, Moscow, 1961, which contains some additional translations of texts not otherwise available;

J. J. M. de Groot, *Die Hunnen der vorchristlichen Zeit*, Berlin and Leipzig, 1921;

E. Chavannes, *Documents sur les Tou-kiue (Turcs) Occidentaux*, St. Petersburg, 1900; reprinted anastatically with additional matter by A. Maisonneuve, Paris, no date (cited as Chavannes *op. cit.*);

Liu Mau-tsai, *Die Chinesischen Nachrichten zur Geschichte der Ost-Türken (T'u-küe)*, Wiesbaden, 1958 (cited as Liu *op. cit.*);

J. R. Hamilton, *Les Ouigours à l'époque des cinq dynasties*, Paris, 1955 (cited as Hamilton *op. cit.*).

I have also consulted W. Eberhard's book cited above, and in summarizing the course of events down to about the fifth century A.D. I have got a great deal of help from two books which contain little in the way of translation but rather summarize and interpret the original authorities. These are:—

W. M. McGovern, *The early empires of Central Asia*, Chapel Hill, 1939;

A. N. Gumilev, *Khunnu*, Moscow, 1960.

It has, I know, been said that McGovern's book contains mistakes and is written in an unscholarly fashion. It certainly does contain mistakes, but it is not alone in that; if it contained none it would be unique. It is also

true that the author has an irritatingly " bright " style, uses an unsightly system of transcription and tucks his references away in notes at the end where they are liable to be overlooked; but the unscholarliness is much more apparent than real, and the book gives a much livelier and more coherent account of the subject than more scholarly books like de Groot's, which incidentally uses an even more unsightly system of transcription.

The great merit of Gumilev's book is that it attempts to coordinate the statements of the Chinese historians with the data provided by archaeological investigation. It is perhaps too early to decide whether all his conclusions are acceptable; in particular, some of the conclusions which he quoted from G. E. Grumm-Grzhimaylo's *Zapadnaya Mongoliya i Uryankhayskiy Kray*, Leningrad, 1926, seem to me to be open to question, but at any rate a start has been made in this field.

If we examine the accounts in the Chinese authorities of the seventh century A.D. and a little later of the tribes who were, we know, at that time talking Turkish, it is clear that the prevailing theory then was that they were all " descended " from a common ancestor, the Hsiung-nu. For example the *Chou Shu*, which was finished in about A.D. 636, says (Liu *op. cit.* p. 5) that " the Türkü are a particular tribe of the Hsiung-nu; the family name is A-shih-na." Similarly the *Sui Shu*, which was finished at about the same time says (Liu *op. cit.* p. 127) that " the ancestors of the T'ieh-lê were descendants of the Hsiung-nu." The *Chiu T'ang Shu*, which was finished in about the middle of the tenth century but largely based on earlier material, makes the same statement (Chavannes *op. cit.* p. 87) about the Uyğur. It is true that other theories were also current at this period. The *Chou Shu* in the same memorandum as that quoted above says (Liu *op. cit.* p. 5) that " according to another tradition the ancestors of the Türkü came from the So country, which lay north of the Hsiung-nu," while the *Sui Shu* says (Liu *op. cit.* p. 40) that " the ancestors of the Türkü were mixed barbarians (*hu*) of

P'ing-liang (in Kansu), with the family name of A-shih-na." This is not incompatible with the first statement in the *Chou Shu*. Both agree that A-shih-na was the family name, presumably of the ruling clan, of the Türkü, and Hsiung-nu might well at this period have been described as "mixed barbarians." The plain fact is that the Chinese authorities of that period were rather uncertain about the earlier history of these tribes, but the favourite theory was that they were descended from the Hsiung-nu. We do not know exactly what ethnic connotation the Chinese at this period attached to the term Hsiung-nu, but there were certainly at this time clans who claimed to be by origin Hsiung-nu and probably spoke Turkish (see Eberhard *op. cit.* p. 309 *etc.*).

With this we can couple the fact that the *Chin Shu* which was written in the middle of the seventh century, but dealt with the period from A.D. 265 to 419, quotes the text in Chinese transcription of a distich allegedly in the Hsiung-nu language datable to the early fourth century.* None of the explanations produced so far are wholly satisfactory, but the language is clearly some kind of Turkish.

The evidence assembled above is sufficient to present a *prima facie* case for the hypothesis that the Hsiung-nu were a Turkish people and spoke an earlier form of the Turkish language, but to turn this hypothesis into a fact it will be necessary to prove that the earlier Hsiung-nu names and words of which a good many have survived in the Chinese records can be explained as Turkish, or at any rate look more like what Turkish might have been at this remote period than anything else. Until Prof. Pulleyblank's paper has appeared and been fully digested it would be premature to express a firm opinion on this point, but one or two tentative steps have been taken in this direction. In *Turk, Mongol, Tungus* I put forward the theory that the Chinese scription of the title of the supreme Hsiung-nu ruler in the *Ch'ien Han Shu, ch'êng-li ku-t'u shan-yü*, represented the Turkish words **teŋri: kutu: *d̲avğu:**. Prof. Pulley-

* See J. Benzing, *Das Hunnische, Donaubolgarische und Volgabolgarische* in *Philologiae Turcicae Fundamenta, Wiesbaden*, 1959, page 686.

blank has satisfied me that I was wrong about the last word, but there is no doubt at all that *ch'êng-li*, which is specifically translated "heaven," does represent an earlier form of the eighth century Turkish word **teŋri:** "heaven." Prof. Pulleyblank has also firmly established the identity between one or two Hsiung-nu titles and titles in early Turkish. Unhappily all the words which have so far been firmly identified as existing in both languages, including the titles just mentioned, are the kind of words which might pass from one language to the other as loan words, just as **teŋri:** later passed from Turkish to Mongolian. Moreover, even admitting that the fourth century distich mentioned above was both Turkish in language and Hsiung-nu in origin, it cannot be confidently asserted that "Hsiung-nu" had the same ethnic content at this date as it had five hundred years earlier. Pending further confirmation, therefore, the hypothesis that the Hsiung-nu were and always had been Turks remains a hypothesis, but it is fair to say that it is a more attractive hypothesis than any other. One alternative theory which at one time commanded wide support, that the Hsiung-nu were Mongols, can now I think definitely be ruled out. I hope to show in Chapter XI that the original habitat of the Mongols was the forests of eastern Siberia and Manchuria and that the whole of their vocabulary for the animals and plants of the steppes was borrowed from the Turks. It is therefore chronologically impossible that the Hsiung-nu, who had lived in the steppes from time immemorial, should have been Mongols. Similarly there is nothing at all to suggest that the Hsiung-nu were a tribe of Indo-European stock and spoke a language of the same kind as their western neighbours in the steppes, or a tribe of Tibetan stock and spoke a language of the same kind as their south-western neighbours. The final possibility which has yet to be disproved is that they were ethnically isolated and spoke a language unrelated to any of those mentioned above, of which no other trace has survived.

Whatever the facts about the ethnic character of the

original Hsiung-nu and the language which they spoke, it can at any rate now be taken as certain that the *name* " Hsiung-nu " is identical with the *name* " Hun " which was familiar from India right across to Europe in the first five centuries of the Christian era. W. B. Henning in his paper *The date of the ancient Sogdian letters*, Bulletin of the School of Oriental and African Studies XII, pages 601 ff., showed that in one of the letters describing events in about A.D. 313 the people to whom the Chinese referred to as Hsiung-nu in describing the same events are referred to as *xwn* " Huns." I shall therefore in future substitute the name " Huns " for " Hsiung-nu."

The use of the term " Turkish " in connection with the Huns and their language is of course anachronistic, since it is derived from a tribal name, Türkü, of which there is no trace before the sixth century A.D., but it is hallowed by tradition and more convenient than any other. I have eschewed the unsightly neologism " Turkic " for the language, if only because it ought logically to be spelt either " Turcic " or " Turkik," both of which look grotesque in English. I have already used, and shall continue to use, the term " tribe " in relation to Turkish peoples. It is the most convenient one to use in English but requires some interpretation. Turkish tribal organization and nomenclature are subjects of appalling complexity on which I touched in *À propos du manuscrit Pelliot Tibétain* 1283. It is perhaps best approached by comparing the steppes in which the Turks lived to a nursery floor covered with a lot of toy bricks. Each of these bricks can be taken to represent a clan or " extended family." The clan seems to have been called by the name of the supposed ancestor who founded it, but when a clan expanded and broke up into several clans each of these took a name of its own, and the original name of the nuclear clan became a collective name for the whole collection of related clans, that is the tribe, and not for any individual unit in it. The Turkish word for " clan " was **bo:d̲**, and for a " tribe " **bod̲un**, an archaic plural of **bo:d̲**.

In the earliest period the population was very sparse, and there was no form of political organization; but as time went on and the primaeval clans became grouped in tribes, one of the constituent clans was recognized as the royal clan and, to revert to our simile, instead of the floor being covered with scattered bricks the bricks came to be arranged in little groups or heaps with one brick, representing the royal clan, on the top of the others. Later still one of these tribes subdued some of the others and the little groups were put together into a larger group, still with one brick on the top of the others in each unit and one brick on the top of all the rest. Such organizations were essentially unstable, and at any time the groups might be broken up and the constituent bricks, with or without others, rearranged in new groups with the same or different bricks on the top of each unit and a different brick on the top of the whole group. The clan/brick which was on the top of the old heap might be destroyed or might survive in a less prominent position retaining its old name. Thus for example the name Hsiung-nu (Hun) survived as an ethnic term in its original form (see Eberhard *op. cit.*), long after the last Hsiung-nu "empire" or "kingdom" had been destroyed, and probably later reappeared as *Hun*, the name in Chinese transcription of one of the constituent tribes of the Tokkuz Oğuz confederation in the sixth century A.D., and as *Kun*, the name of a Turkish tribe known to Arabic geographers in the eleventh century. It should be added that a clan which had at one time been a royal clan usually retained a certain prestige and subsequently appeared as the royal clan of a later constituted tribe or confederation. It also appears that clans which had at one time been incorporated in a particular tribe sometimes gave up their original clan name and assumed the name of the tribe in which they had been incorporated. This explains why some of the sub-tribes of the modern Kırğız bear ancient tribal names like Kıpçak and Tölis which were quite unconnected with the ancient Kırğız.*

* See *Trudy Kirgizskoy Arkheologo-Etnograficheskoy Ekspeditsii I*, Moscow, 1956, map facing page 168.

The Chinese authorities are full of lists of Turkish tribes which were constituent parts of various confederations at various dates. Analyses of some of these lists will be found in Chavannes *op. cit.*, Liu *op. cit.* and Hamilton *op. cit.* and a detailed study of the Tavğaç (T'o-pa) tribes in Eberhard *op. cit.*, Chapter 23. Careful etymological investigation of some of these lists has been carried out, but only a rather small minority of the names mentioned can be identified with names appearing in Turkish or other non-Chinese authorities. This is partly due to the fact that, as I pointed out above, the phonetic values of the Chinese characters concerned in the various authorities cited have still to be established, so that it is often difficult to determine whether a scription in one list is intended to represent the name of the same Turkish tribe as a fairly similar scription in another list, or whether the two scriptions are intended to represent different names. Moreover Chinese historians had a habit of building up monographs on particular subjects by a scissors and paste technique, and incorporating in them passages quoted without acknowledgement from earlier authorities. This may well result in different Chinese scriptions of the name of a particular tribe appearing side by side as if they were intended to represent the names of different tribes. It is, however, very probable, considering the paucity of Turkish tribal names in non-Chinese authorities prior to, say, the thirteenth century, that the Turkish equivalents of many of these Chinese scriptions will never be discovered.

The Chinese were primarily interested in their external political relations with regions beyond their borders, and although they sometimes compiled memoranda of great interest on the peoples inhabiting them, they were not greatly interested in the internal affairs of those areas, except to the extent that they could use dissident elements within a particular " kingdom " to weaken rulers who were getting uncomfortably strong. They were therefore in the habit of saying that a particular tribe or confederation occupied a particular area at some stated time as if this

implied that there had been a complete change of population in that area, whereas except in a few isolated cases all that had happened was that political control had passed from one tribe to another, accompanied perhaps by some similar changes in the control of the smaller constituent parts; in our previous simile the groups of bricks had been shuffled and rearranged with one or two new bricks on the top of them. Thus, for example, when control of the steppes passed from the Huns to the Hsien-pei, from them to the Juan-juan and from them to the Türkü, it is probable that the actual change of population was small; on each occasion probably one or two clans or tribes moved in from one direction, and one or two moved out in another, and the main bulk of the population remained what it was before, but under a new master. Having regard to all these facts, it is extremely difficult to get a comprehensive picture of the overall position in the steppes at any one time, or a continuous history of more than very few of the tribes which inhabited them, even though we can get from the Chinese records very detailed accounts of the history of individual Turkish tribes and their rulers for limited periods.

The origin of the Turks is veiled in the mists of prehistory. If we wish to approach the problem through purely Turkish material, then obviously the only possible starting point is an analysis of the basic Turkish vocabulary in the earliest form accessible to us. I cannot pretend to have undertaken any elaborate analysis of that vocabulary comparable to the analysis of the Mongolian vocabulary in Chapter XI, but my general impression is that it was rich in terms for the flora, fauna and natural features of the steppes, but weak in forest terminology. It was therefore almost the exact converse of the basic Mongolian vocabulary as it is analyzed in Chapter XI. The Turks were in a position to lend the Mongols a good supply of the technical terms (animal and crop names and the like) of animal husbandry and agriculture. It seems reasonable therefore to infer that at about the time the first loans were made

the Turks were primarily a pastoral people with some interest in agriculture. Most of these technical terms look like native words, but one or two loan words have been identified, and more may later be found, which enable us to see that the Turks must have had contacts with neighbouring peoples at a very early date, certainly preceding their contacts with the Mongols. It can hardly be denied that **öküz** " an ox," which was certainly a Turkish word in about the fifth century A.D. and probably a good deal earlier, is a loan word from some Indo-European language, probably taken from *okso* " ox " in " Tokharian B," the language which was still current in Kucha in the second half of the first millennium A.D. The word suggests that the Turks had learnt at any rate some of their animal husbandry from Indo-Europeans, though no doubt much earlier than this. That this contact existed is confirmed by the Turkish eighth century word **tümen** " ten thousand " which is clearly a loan word from " Tokharian B " *tumane*. The use of loan words for the higher numerals is a common phenomenon among primitive peoples; the stock example is the Basque word for " a thousand," *milla*, which was borrowed from Latin to supplement the native numerals, which did not go above a hundred.

There is also good evidence that the Turks were in contact with their southern neighbours the Chinese at an early date. Apart from Chinese titles like **çigşi:, tutuk** and **seŋün,** and words like **kunçu:y** " spouse, consort," which may not go back very much earlier than the eighth century, there is at least one word **beg** " the head of a clan," which, if it really is a Chinese loan word, must be a great deal older, since it is associated with the very earliest stages of social/political organization. It seems to reproduce the Chinese word *po*, " the head of a hundred men " (No. 9,348 in H. A. Giles, *Chinese-English Dictionary*, London, 1912; *pǎk* in Karlgren's " Ancient Chinese," *big/pig* in the eighth (?) century Tibetan transcriptions).

There is also some etymological evidence of early Turkish contact with the Saka, who were established in

south-western Sinkiang in the early centuries of the Christian era, but perhaps not much earlier. The Turkish word **ton**, "garment," noted in the eighth century, is believed to be derived from Saka *ṭauna*, "clothing," but it is not suggested that the Turks wore no clothes until they met the Saka.

Thus the ancestors of the historical Turks must have been in contact at a very early period with Chinese to their south and Indo-Europeans to their west, and, perhaps rather later, with Sakas to the south-west, and must at this time have been a nomadic pastoral people living somewhere in the steppes. Both their general geographical location and their way of life were therefore indistinguishable from those of the Huns as described in the Chinese records. In these circumstances it seems not unreasonable to start our history of the Turkish peoples with the Huns, even though there is no solid proof that the Huns were at this period Turkish by race and language.

When the Huns first appeared in history they were described by the Chinese as a tribe "wandering about in search of pastures and water," that is they were a nomadic people living somewhere north of the Great Wall of China in the belt of steppes which runs from the Manchurian forests in the east through Inner and Outer Mongolia to Zungaria in the west. They maintained themselves primarily by animal husbandry, combined with hunting and food-gathering and perhaps a little primitive agriculture. This was supplemented when times were hard, and sometimes perhaps merely because life was dull, by raiding the adjacent settled areas to the south, that is China and the oases of Sinkiang (Chinese Turkestan). When the governments of these settled areas were sufficiently organized to be susceptible to pressure, but not strong enough to resist that pressure, the Huns also practised a certain amount of polite, or not so polite, blackmail, the terms of which were that they would not raid the settled areas if the governments concerned would make it worth their while not to do so. It was to resist pressure of this kind that various lengths of

the Great Wall were built at different dates from about the middle of the fourth century B.C. onwards.

Earlier than this stage there must have been a pre-pastoral stage in which the earliest Huns, amounting probably to no more than a few scattered clans without any developed tribal organization, maintained themselves simply by hunting and food-gathering. Only a limited amount of archaeological investigation has been carried out in the steppes, and in any event the archaeologists very sensibly refuse to identify the archaeological remains of this very early period with specific peoples or, *a fortiori*, peoples speaking a particular language. There is, however, amply evidence that there was some kind of population in the steppes in the second millennium B.C., or even earlier, and that these steppe peoples were in contact with settled populations at both ends of the steppes.

The question of the earliest intercourse between China and the West is still hotly debated; evidence is still being accumulated and theories formulated and revised, but some points seem now to be established. Although it has now been proved fairly conclusively (see Chêng Tê-k'un, *Archaeology in China, Vol. II, Shang China*, Cambridge, 1960, pages 156 foll.) that bronze casting was evolved independently in China and not introduced from the West, there is a good deal of evidence of one sort or another that there were cultural contacts between China and the West at a very early date, and that the principal cultural exchanges which did take place proceeded *via* the steppes. There is also, it is true, evidence (see V. E. Larichev, *Bronzovyi vek severo-vostochnogo Kitaya*, Sovetskaya Arkheologiya, 1961, Part I) that there was some cultural contact in the Bronze Age between the inhabitants of north-east China and the forest peoples of Manchuria, Trans-Baikalia and even further west, but these contacts seem to have been quite distinct and not to have affected the rest of China. The cultural exchanges were not, of course, only in one direction. In the Minussinsk basin in southern Siberia beyond the mountains to the north of the steppes some of the bronze

НАУЧНАЯ БИБЛИОТЕКА

weapons and implements of the Karasuk stage, round about 1000 B.C. plus or minus one or two centuries, are indisputably of Chinese design, if not actually of Chinese manufacture, and these must have crossed the steppes from the east. Thus throughout the second millennium B.C. there must have been economic and cultural exchanges to and fro across the steppes, and this must surely mean that the nomadic steppe peoples acted as middlemen. As it was the Huns that in the last few centuries before the Christian era were the inhabitants of those parts of the steppes which were immediately adjacent to China it seems reasonable to suggest that the earlier inhabitants of this part of the steppes were also Huns.

There is evidence that Indo-European tribes had reached Sinkiang and the western section of the steppes at a very early date, probably some time during the second millennium B.C. These include the ancestors of the inhabitants of the oases of Kucha and Karashar in Sinkiang who were still speaking "Tokharian" in the first millennium A.D., and the ancestors of the peoples whom the Chinese called Wu-sun and Yüeh-chih. There has been a great deal of discussion about the probable ethnic status of these two peoples. The most popular theory is that they spoke some kind of Iranian, but it is perhaps more probable that they too spoke "Tokharian." Similarly there were almost certainly as early as the first millennium B.C. tribes speaking some kind of Tibetan language in the more mountainous country to the south of Sinkiang and the adjacent steppes in what later became northern Tibet and parts of Kansu and Szechwan. It is, however, quite impossible to delineate at all confidently the ethnic frontiers between these various tribes at this early period. Equally it is impossible to state confidently that there were not in this area some tribes, perhaps even the Hsiung-nu themselves, who originally spoke a language different from all those already mentioned which has completely disappeared leaving no trace of its existence, except perhaps a few words in Chinese transcription.

The earliest references to the Huns in the Chinese records purport to relate to events which occurred in the early part of the second millennium B.C., and obviously contain legendary and improbable elements. There are intermittent references to peoples who can plausibly be identified with the Huns from then onwards. The information begins to be voluminous and precise when it relates to events which occurred in the last two decades of the third century B.C. At that time a man whom the Chinese called T'ou-man for the first time organized the Hunnish tribes into a powerful confederation and became the first Hunnish supreme ruler or *shan-yü*. At that time the Huns seem to have occupied a section of the steppes immediately north of the Chinese frontier between, say, 100°E and 115°E, with its centre of gravity more or less due north of the great (Ordos) bend of the Yellow River. It is quite uncertain how far north T'ou-man's dominions extended, perhaps no further than the Gobi desert, which forms a sort of natural frontier between southern, or Inner, and northern, or Outer, Mongolia. The country occupied by the Huns was adjacent on the west to that occupied by the Wu-sun and Yüeh-chih and on the east to that occupied by peoples whom the Chinese called Tung Hu, "eastern barbarians."* North of the Huns were various tribes which, as we shall see below, were almost certainly Turkish.

T'ou-man's successor was called by the Chinese Mao-tun.†

* I pointed out in *Turk, Mongol, Tungus* that a great deal of needless confusion has been caused by transcribing Tung Hu as "Tungus." In fact, whatever the Tung Hu were ethnically they were certainly not Tungus.

† The archaic Chinese pronunciation of this name was *môg-twən*, and it has been customary to equate it with the Turkish name **Bağa:tu:r**. This cannot be taken as proved, but in any case **Bağa:tu:r** was a very old Turkish name. In the form *Mo-ho-tu* (Giles, Nos. 7,977, 3,994, 12,087; "Ancient Chinese" *muo-γâ-tuət*) it constantly appears in the Chinese records as one of the names of many rulers both of the Eastern and of the Western Türkü from the late sixth century onwards (see Chavannes *op. cit.*). **Bağatur** is given in an Arabic authority as the name of the (Turkish) Xākān of the Khazars in the middle of the eighth century, see K. Czeglédy, *Khazar raids in Transcaucasia in A.D.* 762-764, Acta Orientalia Hungarica, XI, 1-3, 1960. The signatory of the letter in Runic script discovered by Sir Aurel Stein at Tun-huang and first published by V. Thomsen in J.R.A.S., 1912, pages 218 ff., calls himself Bağa:tu:r Çigşi:, the latter the normal Turkish scription of the Chinese title *tz'ŭ-shih* (Giles, Nos. 12,412, 9,893; "Ancient Chinese" *ts'iäk ʂi*) "district magistrate." Although in all these instances it is used only as a proper name, the

Mao-tun, who reigned from about 209 to 174 B.C., greatly enlarged the Hunnish dominions. He first "destroyed" the Tung Hu, no doubt incorporating many of them in his own dominions; to return to our earlier simile he knocked down the Tung Hu heap of bricks and incorporated most of them in his own heap. The only Tung Hu tribes who are said to have escaped were the Wu-huan who established themselves in eastern Inner Mongolia, just north of the Great Wall, and the Hsien-pei, who got as far away from the Huns as possible into the extreme north-eastern corner of the steppes in eastern Outer Mongolia and western Manchuria.* Next he struck north and conquered five other tribes. The exact meaning of "north" in this context is uncertain, it may have meant no more than "north of the Gobi." The Chinese scriptions for the names of these tribes were:—*hun-yü*, *ch'ü-shê*, *ting-ling*, *ko-k'un*, *hsin-li*. Nothing has yet been made of three of these, but the other two undoubtedly have Turkish connexions. A good many theories have been put forward about the Ting-ling, but Prof. Pulleyblank tells me that he had very little doubt that it is an earlier scription of the name later represented by *t'ieh-lê*. The name itself, perhaps

fact that in the earliest period it was usually only one of a string of names or titles suggests that it was originally a common noun or adjective, like many Turkish names. In any event it became a loan word in Mongolian from the thirteenth century (the Secret History) onwards in the form *baǧatur/ba'atur* not as a proper name but as a common noun meaning "hero, outstanding warrior." It reappeared in Turkish, but as a common noun not a proper name, in the fourteenth century *Nabcu'l-farādīs* and the *Codex Cumanicus* in the spelling *bahatur/bahadur*, which proves it to have been a re-introduction from Mongolian. It has ever since the Mongolian invasion been a common word in Persian and still survives in India as the title *Bahadur*, a legacy from the Mogul Empire. The alternative theory that it was originally Mongolian and always a loan word in Turkish seems to be excluded by the fact that it was used as a proper name by the Western Türkü as early as the sixth century.

* In *Turk, Mongol, Tungus* I suggested that *wu-huan* was a Chinese scription for the well-known Turkish tribal name **Oğuz**, but Prof. Pulleyblank tells me that this is improbable. This deprives me of the most convincing argument in favour of the hypothesis that the Tung Hu were Turkish tribes. Nevertheless this still remains the most plausible hypothesis, since the T'o-pa (Tavğaç) who emerged from the Hsien-pei, as will be shown below, unquestionably spoke a form of Turkish. No satisfactory non-Chinese equivalent of the scription *hsien-pei* has yet been discovered, but it is possible that *Saviroi*, the name of a Hunnish tribe in south-eastern Europe mentioned by Byzantine authorities of the fifth and sixth centuries, and *Suwār*, the name of a tribe associated with the Volga Bulgars in the eleventh century which is mentioned by Kāşğarī, were later scriptions of the same Turkish name as *hsien-pei*.

Tiglig* or **Tigrig*, has not so far been found in any non-Chinese authority, but there is no doubt that the T'ieh-lê were a Turkish confederation, and it seems reasonable to suppose that the Ting-ling were also. It has long been agreed that *ko-k'un* and other similar scriptions represent the name **Kırğız. The Kırğız played a prominent part in history in the eighth and ninth centuries and have given their name to one of the Associated Republics of the Soviet Union. Thus we can feel reasonable confidence that even if the Hsiung-nu were not Turks there were Turks in the steppes in the third century B.C. By the seventh century A.D. the Kırğız seem to have been living in Tuva north of the Tannu-ola Mountains. It is possible that it was Mao-tun who drove them out of the steppes into this area.

Next Mao-tun drove the Indo-European tribes further west and imposed his suzerainty on them and on the Indo-European oasis states in Sinkiang. Finally he attacked China with inconclusive results, leading to the establishment of treaty relations and a periodical exchange of gifts, which amounted in effect to the Chinese payment of blackmail as the price of security from raids.

His successor, whom the Chinese called Ki-yü, is chiefly remembered for the fact that he drove the Yüeh-chih right out of the steppes and so started a chain reaction leading to the destruction of the Graeco-Bactrian kingdom. This is probably one of the few cases in which an ethnic unit moved as a whole. It seems likely that the whole nomadic Indo-European population of the steppes moved out in one direction or another and left the whole of the steppes to the Huns and their nomadic vassals, although the Indo-European inhabitants of Sinkiang still retained their hold on the oasis states as Hunnish vassals.

It is unnecessary to pursue the history of the Huns in great detail beyond this point. Briefly the Hunnish Empire started to decline in about 140 B.C., as the strength of China grew, and in the first quarter of the first century B.C. its centre of gravity gradually started to move westwards. At about the same time the Wu-huan threw off Hunnish

suzerainty and established their independence in eastern Inner Mongolia, and the Hsien-pei, if they had ever accepted Hunnish suzerainty, which seems doubtful, ceased to do so. During the next few years the Huns suffered a series of reverses, and in about 55 B.C. a member of the royal house called by the Chinese Chih-chih (a brother of the man called by the Chinese Hu-han-hsieh who had just succeeded in establishing himself as Shan-yü) revolted against his brother and also proclaimed himself as Shan-yü. The "empire" split into two. The Southern Kingdom under Hu-han-hsieh in Inner Mongolia became a vassal state subject to Chinese suzerainty, and the Northern Kingdom under Chih-chih retained its independence. Chih-chih, a man of great energy and enterprise, enlarged his kingdom to the west and for the first time crossed the mountains to the west of the Zungarian steppes into what is now Kazakhstan. He engaged in a series of complicated local wars, sometimes in alliance with one of the local, no doubt Iranian, kingdoms and sometimes with another, and even built himself a capital in the area, probably on the River Talas. The Southern Kingdom took advantage of his preoccupations in the west to edge northwards, and, when Chih-chih was finally killed in 36 B.C. by the combined efforts of the local people and a Chinese army, assumed control of all the steppes, except probably the eastern extremity, but still as a vassal kingdom. It is uncertain whether when Chih-chih was killed and his kingdom overthrown all the Huns in Kazakhstan were exterminated. A. N. Gumilev in his latest article on the subject, *Talasskaya bitva 36 g. do n.e.*, in *Issledovaniya po istorii kultury narodov vostoka*, Moscow-Leningrad, 1960, comes to the conclusion that they were, but it seems possible that a nucleus remained and built itself up with the arrival of more groups of Huns from the east during the next two or three centuries until it finally formed the confederation which invaded Europe in the fourth century A.D. Be that as it may, we hear no more of these Western Huns until that date.

In the east the collapse of the Former Han dynasty, and

the accession to the throne of the usurper Wang Mang in A.D. 9, created a new state of affairs. In the confusion which ensued the Huns threw off Chinese suzerainty and invaded north China in cooperation with the Wu-huan. After a time an uneasy peace was restored, but without any suggestion of Chinese suzerainty. The Huns also took this opportunity to re-establish their suzerainty over the Wu-huan and Hsien-pei.

By the middle of the first century, however, the situation had changed again. China under the Latter Han dynasty regained its strength, and as a result of internal discord the Huns again split into Northern and Southern Kingdoms, the latter ruling Inner Mongolia under Chinese suzerainty, the former Outer Mongolia and some adjacent areas as an independent state. The two were almost continuously in a state of mutual hostility and the Wu-huan and Hsien-pei took advantage of this to throw off the Hunnish yoke. The Chinese persuaded the Wu-huan to settle immediately north of the Chinese frontier and accept their suzerainty, but with the Hsien-pei, who were not an organized "kingdom," but merely a very loose confederation, if indeed that, they succeeded only in establishing some kind of diplomatic relations and a rather one-sided exchange of gifts, the Hsien-pei delivering severed Northern Hunnish heads in return for titles and valuable merchandise of various kinds.

The result of all this was that while the Southern Kingdom remained fairly solidly in control of Inner Mongolia, less the area occupied by the Wu-huan in the east, and served as a useful buffer for China, the Northern Kingdom was edged towards the west by the Hsien-pei, its only common frontier with China being in the area of, and adjacent to, Sinkiang. The struggle between China and the Northern Kingdom for the control of this area proceeded almost continuously throughout the last quarter of the first and the first quarter of the second century A.D. without producing any lasting settlement. The Hsien-pei took advantage of the preoccupations of the Northern Huns

elsewhere to continue their attacks on them and in A.D. 87 even succeeded in capturing and skinning the then reigning Shan-yü. By the end of the first century they had assumed control over almost the whole of Outer Mongolia, confining the Northern Huns to the western extremity of the steppes and Zungaria. It is very probable that at about this time some sections of the Huns moved over the mountains into Kazakhstan either to join the remains of Chih-chih's people, if they still survived, or as small clans or tribal units on their own. As the Hsien-pei drove the Northern Hunnish government westwards they no doubt assumed control of a number of tribes which had hitherto regarded themselves as Huns. To return to our earlier simile, they pushed over part of the Northern Hunnish group of bricks and established a rather disorganized set of groups of their own.

Early in the second century A.D. the Southern Huns took advantage of Chinese preoccupations further west and at home to reassert their independence, and the Wu-huan joined them, but this rebellion soon collapsed. Ten years later the Northern Huns also became active again and resumed hostilities in Sinkiang, but neither side was able to achieve a lasting success, and things drifted on in much the same way until about the middle of the century. During this period the Chinese hold over the Southern Kingdom and the Wu-huan continued to be precarious. From about A.D. 115 onwards the Hsien-pei too began to launch a series of attacks on north-eastern China, which continued for about twenty years and then lapsed into quiescence.

In about A.D. 150 a man whom the Chinese called Tan-shih-huai assumed control of the Hsien-pei and for the first time organized them as a powerful steppe "kingdom." Apart from constituting a standing threat to China he extended his dominions west and north of the steppes and conquered and controlled all the territory formerly ruled by the Northern Huns. He is also remarkable for having taught his people to catch and eat lake and river fish. However he died in A.D. 180 and the "kingdom" created by him almost immediately disintegrated.

The Southern Kingdom disintegrated a few years later and so the whole of the steppes was occupied by small tribes managing their own affairs without any central organization. In our earlier simile all the big heaps had fallen down and all that was left was a few small ones. It is very possible that if the Huns had originally spoken a language of their own different from Turkish, it was at about this time that it died out, since after this period they were no more than "mixed barbarians," living in a milieu which at any rate by the sixth century was almost exclusively Turkish-speaking.

Formally the Southern Kingdom came to an end in A.D. 216, and in A.D. 220 the Latter Han dynasty came to an end also, and China disintegrated into three separate and rival kingdoms. Almost the whole of north China constituted the Kingdom of Wei, the largest and strongest of the three. It was strong enough to ensure that it was not invaded across its northern frontiers, but not strong enough to exercise any control beyond them. At about this time a number of Southern Huns submitted and entered China and were organized in a number of small groups by the Wei government, each unit under its own headman and the whole under a chief headman. Beyond the frontier the political vacuum was gradually filled by the Hsien-pei who spread their influence south into Inner Mongolia. It was at this time that the Tavğaç (in Chinese scription *t'o-pa*) clan first assumed the leadership of the whole Hsien-pei confederation.

In about A.D. 265 the Wei and the other two kingdoms collapsed and China was nominally reunited under the Chin dynasty. But the central governmental machine was too weak to govern the whole of China effectively and by the end of the century civil war was endemic throughout the whole of north China, and small kingdoms, the rulers of which often called themselves "Emperors," began to appear and disappear with monotonous regularity. Almost none of these self-styled Emperors were native Chinese; one or two were Tibetans or Mu-jung (Hsien-pei), but the

great majority were Huns, descendants of the Huns who had been settled in north China under the Wei dynasty.

The Tavğaç, who had strengthened their grip over Inner Mongolia by about A.D. 260, continued their pressure southwards and by A.D. 310 assumed control of some territory south of the Great Wall. However in A.D. 316 civil war broke out in the confederation and the pressure eased. The Tavğaç ceased to be able to control the whole confederation, and the Mu-jung clan, or tribe, organized a confederation of its own in south-west Manchuria, which in the second half of the century pushed down into north-east China and set up the two Yen dynasties.

Towards the end of the fourth century the Tavğaç revived under a series of energetic rulers, and by degrees assumed control of the whole of north China, ruling as the Northern, or Yüan, Wei dynasty (A.D. 386-535).

The political vacuum which was left in the steppes was soon filled by a people whom the Chinese called Juan-juan, Ju-ju, Jui-jui or Jou-jan. No tribal name has so far been discovered in any non-Chinese authority which could correspond with these scriptions. Their rise to power started in the second half of the fourth century, and their ruler declared himself a supreme ruler with the title *kağan* in A.D. 402, apparently the earliest use of the word, which is an old, perhaps Hunnish, one, in this meaning. Before long the Juan-juan controlled the whole of the steppes from their eastern extremity to Zungaria and were fighting the Wei for the control of Sinkiang. This state of affairs continued for nearly a century and a half, but from about A.D. 520 onwards civil war began to break out between various members of the royal family and the Juan-juan began to lose control. They finally collapsed in the middle of the sixth century. At about that date a Turkish tribe whom the Byzantine authorities called " Avar " appeared in eastern Europe. Chavannes *op. cit.* made out a plausible case for the theory that these Avar were Juan-juan who had fled to the west, but positive proof is still lacking and doubt has since been expressed about the validity of some

of Chavannes' evidence. It is very desirable that this point should be settled one way or the other. It is possible that this could be done by comparing more closely the references to the downfall of the Juan-juan "empire" in Chinese, Byzantine and Moslem authorities since this was an event which affected many countries and governments in Europe and Asia. There are reasons for supposing that the Juan-juan, like the Tavğaç, with whom they had some clan names in common, were by origin a Hsien-pei tribe and talked the same kind of Turkish as the Tavğaç, but this requires further investigation.

It is reasonably certain that the great bulk of the tribes whom they ruled in the steppes had been there a very long time and had been ruled by the Hsien-pei and the Huns before them. It seems probable that when the Juan-juan rose to power these tribes were organized in some kind of a loose confederation called by the Chinese *kao-ch'ê* "tall carts" because their members carried their families and possessions about in such vehicles.

Whatever the ethnic status of the Juan-juan may have been, there is no doubt that it was a Turkish tribe, the Türkü, the eponymous tribe of all the Turkish peoples, who headed the revolt which put an end to the Juan-juan "empire" in about A.D. 552. The name itself does not appear to be known before this date. As stated above, the story favoured by the Chinese authorities was that the royal family of this tribe, called by the Chinese A-shih-na, who at that time were living in the Altai mountains, had originally been one of the small groups of Hunnish refugees who had been settled in north China by the (first) Wei dynasty, and in the fifth century were living in Kansu under the rule of the Hunnish Western Ch'in dynasty. They fled to the Juan-juan with 500 families when the Northern Wei Emperor put an end to that dynasty in A.D. 431, and it seems very likely that they were soon on the move again and fled to Altai to escape the Juan-juan.

This was not of course the first case of a tribe leaving the steppes, the process had begun five centuries earlier

when Chih-chih crossed the mountains to the west, but it seems to be the only recorded instance of a tribe which had left the steppes later returning to them. The early westward movements of Turkish tribes must have included parts of several tribes, for by the fifth century some Oğuz* had moved far enough to the west to have been heard of by the Byzantines under the name of *Ougouroi*, and it is not impossible that the big concentration of Oğuz in the area to the east and south-east of the Aral Sea, who in the ninth century developed some form of organized government under a ruler with the venerable title of yavğu: had begun to gravitate to this area when the Juan-juan were ruling the steppes. This does not of course imply a total migration of the Oğuz to the west. A great many remained behind and played an important part in the history of the seventh and eighth centuries.

When the Türkü returned to the steppes and destroyed the power of the Juan-juan they quickly assumed their mantle and, at the height of their power, controlled not only the whole of Inner and Outer Mongolia but also settled areas far to the west. At the time of their return the main bulk of the population of the steppes seems to have been organized in a loose confederation (perhaps the same as that earlier called by the Chinese *kao-ch'i*) which the Chinese called T'ieh-lê.† The memorandum on this confederation in the *Sui Shu* (translated in Liu *op. cit.* p. 127) does not inspire confidence. It seems to be not so much a list of the confederation itself as a list of all the nomadic

* Since my identification of *wu-huan* with "Oğuz" seems to be an error, this is the earliest occurrence of this name.

† I have pointed out above that this is a later scription for the same name, perhaps **tiglig* or **tigrig* which was earlier transcribed *ting ling*. Almost as much confusion has been caused by transcribing *t'ieh-li* Tölis as was earlier caused by transcribing *tung hu* Tungus. The error was first exposed by Boodberg, but very little examination would have shown that the identification is impossible. The Tölis were never a confederation of any great importance. The name is always mentioned in conjunction with Tardus, the two being the names of the subdivisions of some larger confederation, see E. G. Pulleyblank, *Some remarks on the Toquzoghuz problem*, Ural-Altaische Jahrbücher, XXVIII, 1-2, note on page 55. For example the *T'ang Shu* (see Bichurin *op. cit.*, I, 340) says that the Tölis and Tardus were the two major subdivisions of the Hsieh Yen-t'o, which was itself (*op. cit.*, I, 339) the most powerful component of the T'ieh-lê in the second quarter of the seventh century.

peoples who were subject to the Türkü at the height of their power. Seven geographical areas in which T'ieh-lê tribes lived are enumerated, ranging from " east of Fu-lin (the Byzantine Empire) " to " south of the North Sea (presumably Lake Baikal)." Lists of the tribes in each area are given, each ending " and others." The total number of names mentioned is forty-four, and some of them, for example A-lan (the Alans) are obviously not Turkish. The list of T'ieh-lê tribes in the *T'ang Shu* (see Chavannes *op. cit.* p. 87), though later, seems to be more realistic. It includes only fifteen names, most of them recognizably Turkish. E. R. Pulleyblank (*op. cit.* in the footnote p. 28) has pointed out that one important, indeed at one time perhaps the most important, component of the real Turkish T'ieh-lê was what the Chinese called *chiu hsing* " nine surnames," that is the Tokkuz Oğuz, itself, as its name implies, a confederation of nine tribes. These nine tribes were no doubt located in various parts of the steppes; the best known, and probably one of the most northerly of them, was the Uyğur, whose original habitat according to the Chinese authorities (see Chavannes *op. cit.* p. 89 and Hamilton *op. cit.* p. 2) was the region of the River Selenga.

The Uyğur too were a confederation of ten tribes or clans, of which the Yağlakar were the senior or royal clan.*

At a very early date in its history the Türkü dynasty broke into two parts, commonly called the Western and the Eastern (or Northern) Türkü. The Western dynasty, after passing through various vicissitudes, finally collapsed in the first decade of the eighth century. The Eastern dynasty, also after passing through many vicissitudes, including a period of fifty years of total eclipse in the middle

* To make things more complicated still the Staël-Holstein roll, which is dated A.D. 925, that is nearly two centuries after the collapse of the last Türkü Empire, contains a list of the Tölis and Tardus tribes in which the Yağlakar appear as the royal tribe, or clan, of the Tölis. This is some indication of the complications of Turkish tribal organization which permitted the same tribe, or clans of it, to move from one confederation to another and yet retain royal status. For the Staël-Holstein roll see W. B. Henning, *Argi and the " Tokharians,"* Bulletin of the School of Oriental Studies, IX, 3; the same author's *The name of the Tokharian language*, Asia Major, New Series, I, 2, note on page 162; H. W. Bailey, *The Staël-Holstein miscellany*, ditto, II, 1, 1951; and E. J. Pulleyblank, *The date of the Staël-Holstein roll*, ditto, IV, 1, 1954.

of the seventh century, finally collapsed in A.D. 742. At that date the Uyğur, who had been gradually building up their strength since the beginning of the century, drove down from the north and crushed them. For the main part of our knowledge of the history of the Türkü during these two centuries we are dependent on the Chinese records, but the Orkhon inscriptions (see Chapter V) include funerary memorials of one *kağan* of the dynasty, his brother and two, or perhaps three, of his high officials. The relevant Chinese texts have been translated in two major works, Chavannes *op. cit.* and Liu *op. cit.*, and some of them, in less satisfactory versions, in Bichurin *op. cit.* The history of the Uyğur who crushed the Türkü in A.D. 742 and for a time assumed their mantle as rulers of the steppes, and of the later Uyğur dynasties, has been admirably summarized in Hamilton, *op. cit.*

It is neither necessary for present purposes, nor would it be possible briefly, to summarize the history of the Turkish peoples from the eighth century onwards, but it may be useful to mention a few historical events which have a bearing on the evolution of the Turkish languages. In the steppes Uyğur rule did not last very long and other tribes like the Karluk and the Kırğız assumed a major role. The later Uyğur were more important in the oases of Sinkiang than in the steppes. West of the mountains other tribes like the Türgeş, one of the components of the Western Türkü, were prominent for longer or shorter periods, but in this period the mountains never constituted a fixed political or ethnic frontier, and some tribes controlled territory on both sides of them. In the second quarter of the tenth century a new power emerged in this area. Its tribal basis is obscure; no doubt it included a number of tribes, but the founder of the dynasty was called Kara Xan ("the black *xan*"), and the dynasty itself is known in history as Karakhanid. At the height of its power this dynasty controlled a considerable area and had capitals at Kāşğar in Sinkiang and Balasağun somewhere in the valley of the River Chu. Formally the Karakhanid dynasty ruled

with varying fortunes from A.D. 932 to 1165, but it was of very little significance during the latter part of this period. Not long after the middle of the tenth century its rulers were converted to Islam and from that time onwards more and more Turkish peoples became Moslem, although many of them retained their original paganism. Karakhanid control never extended west of the Aral Sea into the steppes of southern Russia where a succession of pagan Turkish tribes, Peçeneg, Khazar, Kıpçak and so on, were prominent for longer or shorter periods. Not very long after the Karakhanids had seized power, and very probably in order to escape their control, groups of the Oğuz who had been established to the south and east of the Aral Sea began to move southwards. The name Türkmen for these Oğuz sections seems to date from this period, and it has been suggested, though this is perhaps quite wrong, that the name means rather specifically Moslem Oğuz. They first became important when one of their tribal leaders commonly called Selcuk (the spelling dates back to the early unscientific period of mediaeval historiography, the correct form is *Salcuk*) organized them into a great power. Formally the Great Selcuks reigned from A.D. 1038 to 1157, but long before the latter date junior members of the family had established dynasties of their own, and in fact there was a period during the twelfth century when nearly the whole of Moslem Asia was ruled directly or through puppet monarchs by Turks. For example large parts of Asia Minor were ruled by the Selcuks of Rūm from A.D. 1077 to 1300.

The Mongolian invasion early in the thirteenth century created an entirely new situation. Except for Anatolia, Syria and the Arabian peninsula, the thinly populated areas in the north, and parts of the south-east Mongolian rule extended over the whole of Asia and substantial parts of eastern Europe. On the death of Chinggis Xan these vast dominions were parcelled out among his sons and their descendants; but although the rulers and their principal officials, particularly on the military side, were Mongols,

these were so few in number that they never succeeded in imposing their language on their subjects. Enormous damage was done to the economic and cultural life of the conquered peoples, but the invasion had only a small impact on the languages which they spoke, and it was not very long before Turkish revived not only as a medium for the composition of Moslem religious works, but even as the literary language at the Mongolian courts. One curious small concession was however made, a number of these Turkish texts were transcribed in the Mongolian Official Alphabet (see Chapter IX) which the Mongols brought west with them, even though the rulers and high officials for whom they were transcribed had long ceased to speak Mongolian and were probably not ethnically Mongols.

CHAPTER II

THE EVOLUTION OF THE TURKISH LANGUAGES

In Chapter I I have sketched in broad outline the history of the Turkish tribes from the earliest times down to the early middle ages. The next task is to attempt, on the basis of that history, to reconstruct the history of the Turkish languages. Such a reconstruction can only be tentative; I have no means of proving that it is accurate, and perhaps it is not. But there is, so far as I know, nothing inconsistent between this reconstruction and the evidence in our possession, and it will have to do until something better can be produced.

The history of every language, if it could be traced far enough back, would probably be found to have followed much the same course of development. At the dawn of prehistory, when man first became man in the accepted sense of the term, the habitable parts of the world were very thinly inhabited by small groups of human families, associated in hunting and food-gathering bands. These groups, or as I called them in the last chapter clans, lived in complete, or almost complete, isolation from one another, and each of them must have evolved some kind of system of communication between its members, presumably by a range of gestures, grunts, screams and other sounds which were accepted as meaningful by all the members of the group. In this initial stage human methods of communication were presumably much the same as those employed within groups of the higher animals (apes *etc.*) to-day.

At some point in time, not necessarily simultaneously all over the habitable world, the meaningful sounds employed by such an isolated group evolved into a set of what we should probably accept as being words. These words were no doubt at first completely undifferentiated, each one designating a concrete object or an abstract idea. As time

went on these undifferentiated words came, at any rate in most languages, to be used in rather different ways, and crystallized out into various parts of speech, nouns, verbs and so on, though at first no doubt the dividing lines between these classes were very vague and indefinite. At this stage the total vocabulary of the group, that is the words accepted by its members as having specific meanings, was no doubt a rather restricted one, and there must have been a constant need to find words to connote newly discovered things and ideas.

So long as the group existed in isolation these new words could be evolved only in four ways:—

(1) existing words might be given new or additional meanings;
(2) two or more existing words might be combined to convey a meaning more precise than, or even quite different from, the meaning of either word in isolation;
(3) additional sounds might be added to existing words to produce new words, usually with meanings in some way related to the words so elongated;
(4) completely new combinations of sounds might be invented which would be accepted by the whole group as meaningful.

As soon as two groups which had evolved their own languages in this way came into more or less intimate contact with one another, they acquired a fifth source of new words; they could borrow one another's. And in practice experience shows that when one such group acquired a new thing or idea from another, it habitually borrowed the other group's word for it.

At some point in time, sometimes before, but probably nearly always after, contact had been established with some other group, nearly all these groups got so large that they began to split up into sub-groups, and, as contact between the sub-groups became less and less intimate, each began to develop a dialect of its own, differing from the dialects of the other sub-groups by idiosyncrasies of pronunciation or grammar or by the invention or acquisition of new words

not used in the other dialects, or by both. For a time, perhaps a long time, the speakers of any one of these dialects would have had very little difficulty in making themselves understood by the speakers of the other dialects, but in the long run the contacts between the sub-groups weakened to a point at which the dialects became separate languages, the original language, as such, ceased to exist and what came into existence was a family of languages. The circumstances in which a unitary language breaks up into dialects, and these in due course evolve into separate languages, have often been studied, and some influences which promote, and some which inhibit, this kind of fission have been identified. When there is freedom of movement and communication between the clans or tribes which make up a nomadic community, differences are slow to develop. They develop rather quickly when clans or tribes adopt a sedentary form of life, especially when forest-dwelling tribes settle down in separate river valleys with very little cross communication, or when nomadic tribes deliberately part from one another and confine their movements to specific parts of the open country, neither intruding on their neighbours nor permitting intrusion by them.

As soon as an unitary language has evolved into a number of separate languages, each of these becomes a living organism slowly but constantly changing. All the time it is acquiring new words by one of the methods enumerated above. At the same time some existing words are undergoing a subtle change of meaning, and others are falling out of fashion, becoming obsolete and finally disappearing. There are similar slow secular changes in grammar and pronunciation.

It was not much earlier than the nineteenth century that some scholars became interested in the comparative studies of languages. When they did so it soon came to be realized that it could be proved by careful analysis and examination that certain languages were related to one another and descended from a common ancestor, and that what can be called " family trees " of such languages could

be constructed. Thus it has been proved that certain languages, and only those languages, form the Indo-European family, certain others the Semitic family, certain others the Uralian family, and so on.

It has been suggested that the Turkish, Mongolian and Tungus languages form such a family, commonly called the Altaic, and that they are all descended from a lost primaeval language called Altaic or Proto-Altaic. For some years now I have been coming more and more to the opinion that this is an error and that the fact that these languages have a good deal of vocabulary material in common is best explained, not by assuming that they have inherited it from a common ancestor, but by assuming that a prolonged and complicated process of exchanges has taken place between these languages. Whether or not the Mongolian and Tungus languages are genetically related—my own impression, based on quite insufficient knowledge, is that they are not—I am quite convinced that Turkish is not genetically related to either of them. My reasons are stated in Chapter XI.

There is of course a Turkish family of languages, and the Turkish "family tree" can be reconstructed by direct evidence for at least the last twelve hundred years and by indirect evidence for a good deal longer. Indeed it seems possible tentatively to carry the process back to a point at which there was only one Turkish language, the only known descendent of the primitive language of one of the primaeval groups referred to above, and the common ancestor of all the Turkish languages, living and dead, which are known to us.

The Turkish languages can be divided into two main groups distinguished not by differences of vocabulary—even to-day, when later accretions have been stripped off, all the Turkish languages prove to have much the same basic vocabulary (numerals, pronouns and a number of common nouns and verbs)—but by differences of pronunciation, the most obvious of which are that in one group of languages, which we can call for the sake of

convenience the "standard languages," ş and z are common basic sounds, while in the other group, which we can call the "l/r languages," these sounds do not exist as basic sounds but are represented by l and r respectively in words which contain ş and z in standard Turkish. It is clear that the separation into these two groups took place a very long time ago. It has been suggested by the proponents of the Altaic theory that the Turkish l/r languages are "older" than the standard languages and that in the Proto-Altaic language there were no such sounds as ş and z, but that there were two l sounds (l^1 and l^2) and two r sounds (r^1 and r^2), that in Mongolian and some Turkish languages both l^1 and l^2 became l and both r^1 and r^2 became r, while in the remaining Turkish languages l^1 became l, l^2 ş, r^1 r and r^2 z. This rather preposterous theory seems to be based on nothing more solid than the facts that the earliest Turkish loan words in Mongolian were borrowed from an l/r language and that ş and z are not basic sounds in Mongolian. It is in fact not possible to adduce any convincing reason why some "Proto-Altaic" words containing l and r should have passed down both to Turkish and Mongolian with these sounds unchanged, while in other words these sounds changed, l to ş and r to z, in some words passed down to some Turkish languages, while they passed unchanged down to other Turkish languages and Mongolian in these words.

I discussed the l/r languages at some length in *The Turkish elements in 14th Century Mongolian* and *Turk, Moṇgol, Tungus* and shall return to that subject in Chapter XI. The first traces of such a language seem to be found in the Chinese scriptions of native words in the original (Tavğaç) language of the rulers of the Northern (or Yüan) Wei dynasty. These words can be dated to the fifth century. The first layer of Turkish loan words in Mongolian was borrowed from an l/r language, I have suggested this same Tavğaç language, at about the same date. There do not seem to be any other traces of an l/r language in eastern Asia. There is, however, abundant evidence of the early

arrival of a similar language in the southern Russian steppes. There are several names of Turkish tribes in this area in the Byzantine records of the fifth century, and some of these names, for example *Ougouroi* for **Oğuz,** are in the forms which they would have assumed in an **l/r** language. Other examples are *Onogoroi* (**On Oğuz**), *Saragouroi* (**Sara Oğur** < **Sarığ Oğuz**), *Outigouroi*/*Outourgouroi* (**Otur Oğur** < **Ottuz Oğuz**) and *Koutrigouroi*/*Koutourgouroi* (probably a metathesis of **Tokur Oğur** < **Tokkuz Oğuz**). The few remains of the Protobulgar language of the Danubian Bulgars are so obscure that it is difficult to draw any conclusion from them but there are indications that they were in an **l/r** language. It is generally agreed that the earliest Turkish loan words in Magyar were borrowed from some Turkish tribe speaking an **l/r** language (see Chapter V (3)). Similarly there is no doubt that the inscriptions on thirteenth and fourteenth century tomb-stones of the Volga Bulgars which contain words like **hır** (< **kı:z**) " daughter " are in an **l/r** language. The only modern **l/r** language is Chuvash spoken on the middle Volga, which is obviously a direct descendant of the language spoken in that area either by the Bulgar or some related tribe like the *Suwār* mentioned in Chapter I. In *The earliest Turkish loan words in Mongolian* I suggested that the *name* Chuvash is a corruption of **Tavğaç** and there is good authority in the Chuvash language for all the sound changes involved. Initial **ta-** in standard Turkish often becomes **çu-** in Chuvash, as in **ta:ş** " stone " > **t'ul** > **çul;** medial **-ğ-** after another consonant is usually elided, as in **suvğar-** " to drink " > **şăvar-,** and final **-ç** sometimes becomes **-ş** as in **iç** " interior " > **ăş.***

We have no means of ascertaining how an **l/r** language got so far west at this early date. We are in fact almost completely ignorant of what was happening in the first few centuries of the Christian era in the country west of

* For these sound changes see J. Benzing, *Das Tschuwaschische* in *Philologiae Turcicae Fundamenta*, Wiesbaden, 1959. It will be noticed that in Chuvash standard Turkish **ş** has become **l** as in **ta:ş** > **çul,** but that a secondary **ş** has evolved out of standard Turkish **ç** and **s**.

the mountains which enclose the western extremity of the Mongolian steppes and north of the frontiers of the Roman/Byzantine and Sassanian Empires. We know that Chih-chih crossed these mountains in the first century B.C. but did not get very far. There is no reason to suppose that the scattered groups of Huns and other non-Indo-European peoples who crossed them in the next two or three centuries got much farther. It seems to be established, by archaeological evidence rather than written history, that there was a solid block of Iranian peoples, Scythians, Sarmatians and the like, in the steppes of southern Russia and far to the east during this period. When the Huns appeared there in the third quarter of the fourth century they came out of the blue, and the only solid information about them is that although they led a large and miscellaneous horde there were not very many real Huns. Nor has it been satisfactorily proved that the Huns who invaded Europe were ethnically identical with the Huns who, according to the ancient Sogdian letters, were active in China half a century earlier, though it seems reasonable to suppose that they were. The most obvious hypothesis is that some clans or tribes speaking an l/r Turkish formed part of the Hunnish horde or followed closely behind, and if there is anything in the equations Hsien-pei/Saviroi and Tavğaç/Chuvash they may have been descendants of the Hsien-pei who ranged over the Mongolian steppes in the third century A.D., but this is a mere speculation based on no solid facts.

Apart from the languages just mentioned all the Turkish languages known to us belong to the "standard" group. Pending further elucidation of the early Chinese scriptions of Turkish or supposedly Turkish names and words it is a fact that the earliest indirect evidence regarding the l/r languages in the Chinese scriptions of Tavğaç and in the fifth century Byzantine authorities is older than any evidence which we have regarding the "standard" languages, but this is purely fortuitous. An unitary Turkish language of the "standard" type must have existed for a

very long time before it began to break up into dialects and these fell apart into two groups. How this occurred can only be a matter of conjecture, and it is perhaps hardly profitable even to venture a conjecture until we know more of the ethnic and linguistic status of the early tribes mentioned in Chapter I. In *Turk, Mongol, Tungus* I suggested that the original undivided Turkish language was the language of the steppes in the first millennium B.C., spoken alike by the Huns, the Tung Hu and the Turkish tribes to the north (*Tiglig/Tigrig, Kırğız *etc.*), and that this remained the standard language of the steppes, while the l/r language took shape among the Hsien-pei when they fled to the north-east corner of the steppes to escape the Huns at the beginning of the third century B.C. This is, by analogy, the kind of situation in which a language does break up into dialects. In this simple form I can clearly no longer put forward this theory as valid; it has yet to be proved that the Huns at this time spoke Turkish, and the fact that my equation *wu huan*/**Oğuz** cannot be maintained makes it impossible for me to use it as evidence that there was a tribe which is known later to have spoken Turkish already in the neighbourhood of the Chinese frontier in the third century B.C. Nevertheless I think that I can still put forward an alternative and perhaps even more plausible hypothesis. The peoples who spoke forms of " standard " Turkish, certainly in the eighth century and almost certainly at least two centuries earlier, all lived in the central or western steppes and are said to have been " descendants " of tribes who lived in this area a great deal earlier. The Tavğaç, the earliest tribe known to have spoken an l/r Turkish, were " descendants " of the Hsien-pei. It is possible therefore that the earliest l/r language developed its peculiarities, if not when the Hsien-pei cut themselves off from the rest in the third century B.C., at an even earlier period when the Tung Hu were living somewhat isolated from the rest.

Be that as it may, it is reasonably clear that by the fifth century A.D. and probably some centuries earlier there were

two Turkish languages, an early "standard" language, for which it would be more prudent not to suggest a name, since there are so many possibilities, and an l/r language which for want of a better name we can call Hsien-pei or Tavğaç.

By about the middle of the sixth century, when the Türkü returned from their self-imposed exile in the Altai, there had been some changes in the situation. The Tavğaç language seems to have died out in the Far East, chiefly because of the determined efforts of the rulers of the Yüan Wei dynasty to cut themselves off from their "barbarian" past and pass themselves off as Chinese. They were so successful in this that the Türkü had completely forgotten that Tavğaç had ever been the name of a Turkish tribe and used it as the word for "China" and "Chinese" even though the Yüan Wei dynasty had by that time disappeared. In the Far West, however, the principal Turkish language spoken at this time was an l/r one. It is this language that we call in its successive stages Proto-Bulgar, (Volga or Mediaeval) Bulgar and Chuvash. There is no direct evidence that standard Turkish languages had reached this area at that date, but the -z- in their name suggests that at any rate the Khazar, who are mentioned in Byzantine authorities from the early seventh century onwards, did so.

Bulgar is one of the many oddities of Turkish tribal nomenclature. It must originally have been the name of a Turkish tribe living in the Asiatic steppes, but although it appears in the Byzantine records from the late sixth century onwards, no trace of it exists in the Chinese records, except perhaps in a form so distorted that it has not yet been recognized, or in Turkish inscriptions or texts until the eleventh century when some Turks had been converted to Islam and so become part of the Moslem world.

By this time too dialectic differences had begun to appear in the standard languages. The ancestors of the Türkü who took refuge in north China after the collapse of the Hunnish Southern Kingdom at the end of the second century and migrated to the Altai early in the fifth century

must during this period and during their stay in the Altai have been to a large extent isolated from their kinsmen who stayed in the steppes under the control first of the Hsien-pei and then of the Juan-juan. There is therefore nothing surprising about the fact that when they reappeared in the steppes in the middle of the sixth century they spoke a language which was different from, and slightly more archaic in some phonetic and grammatical respects than, the language of their kinsmen who had stayed in the steppes and organized themselves in the *kao ch'ê*/T'ieh-lê confederation. It would in fact have been more surprising if during this period of separation no differences had developed. We have a few inscriptions of the Türkü dynasty which are precisely dated to the eighth century in a language which can conveniently be called " Türkü " and a few manuscripts in the same language (see Chapter V (1) (v)). We have also one or two inscriptions and a great mass of manuscripts, some as old as the eighth century, and one or two perhaps copies of works composed or translated into Turkish as early as the seventh century, which are in a slightly different language. It is the custom to call this language " Uyğur," and this name was in fact used in one or two colophons of texts in this language, although the name " Türk " was more often used in such colophons. The name " Uyğur " is obviously anachronistic, since the language must have been well established in the steppes long before the Uyğur became prominent early in the eighth century, but no useful purpose would be served by suggesting an alternative.

In fact by the eighth century there were actually not two but three slightly different " standard " languages. The phonetic differences between Türkü and the other two, which are discussed in Chapter V (1) (v) mainly affect one consonantal sound and are easily observed. But there are also minor differences of vocalization within the group of Uyğur texts. There are a few Manichaean texts in which the vowels **a** and **e** are habitually used in positions in which other vowels are used in Türkü and in the remaining

Uyğur texts; in a few Buddhist texts similar phenomena occur, at any rate sporadically. The two languages can therefore conveniently be called " Uyğur " and " Uyğur-A " respectively. It was pointed out in Chapter I that the T'ieh-lê confederation, which must have spoken these languages, included a number of tribes scattered over a wide area, and the differences between Uyğur and Uyğur-A are exactly the kind of differences which might emerge between the languages of tribes living in not very close contact with one another. It will be suggested in Chapter V (1) (v) that one possible explanation is that the peculiarities of Uyğur-A were evolved among the true Uyğur while they were still living in semi-isolation in the north before they emerged to " destroy " the Türkü in the middle of the eighth century. If so, it is Uyğur-A that has the better claim to the name " Uyğur," but this is a mere conjecture and may be quite wrong. Be that as it may, Uyğur-A soon disappeared, and Uyğur became the dominant language of the steppes and the Turkish populations of the oases in Sinkiang.

At about this date, or at any rate not much later, a fourth, and phonetically speaking " younger," standard dialect must have evolved in the Far East, the most salient characteristic of which was that in it initial **y-** became **c-**. It was from such a language that the second layer of Turkish loan words was borrowed by Mongolian (see Chapter XI) some time between the eighth and twelfth centuries. It is difficult to find a name for this dialect, but a study of the tribal geography of the eighth century as shown in the Orkhon inscriptions and other slightly later authorities suggests that it may have been the language of the Ottuz Tatar. That tribe was in the far north-east of the steppes at that period, close to the Mongolian tribes who were then beginning to emerge from the forests, and, as will appear shortly, the Tatar were the Turkish tribe best known to the Mongols when they rose to power in the twelfth century. There are several modern Turkish languages which share this phonetic peculiarity, and in the eleventh

century Kāşğarī mentioned it as a distinguishing characteristic of several languages in his day. Both the languages mentioned by Kāşğarī and the modern languages concerned are rather widely distributed, and it is quite possible that this sound change, which is quite common in languages unconnected with the Turkish group, may have occurred independently in different areas, but the modern languages which have initial **c-**, or a similar sound, instead of **y-** include some languages in the North-Eastern (south Siberian) group, which may have inherited it from the early dialect mentioned above, and Kazan Tatar spoken on the middle Volga.

Meanwhile behind their mountain fastnesses to the north the Kırğız continued to speak a language, of which we have some remains at any rate as old as the ninth century, which seems to be very close to Türkü.

Thus by about the eighth or ninth century four or five languages had evolved from the original unitary Turkish language, but there is no reason to suppose that the original vocabulary had disintegrated to any great extent. No doubt some words had been forgotten or changed their meanings in one language or another, and in each language some new words had been invented or acquired, but broadly speaking the old vocabulary had survived intact.

By this period no language existed which had retained in all respects the phonetic structure of the unitary language, but each language accessible to us, either directly or through loan words in other languages (in particular Mongolian and Magyar) had retained some archaic traits which were lost in the rest. By judicious eclecticism, therefore, we can reconstruct the phonetic structure of a very early stage of the language, perhaps even the old unitary language itself, but to avoid begging any questions it would be more prudent to use some neutral term, and I shall call this language simply " pre-eighth century Turkish."

Our knowledge of **l/r** Turkish is so fragmentary and discontinuous that it is better not to attempt to trace its history in detail, remarking merely that the difference

between l/r Turkish and the standard languages was primarily in the pronunciation of certain sounds and probably only to a small extent in matters of word structure, grammar and vocabulary; what is said below about the general structure of standard Turkish is equally applicable to l/r Turkish.

By the eighth century, the date of the earliest continuous Turkish texts which have survived, the language was fully developed and capable of expressing anything that its speakers wished to express. It had an elaborate grammar with a well-developed accidence and syntax. It comprised verbs, nouns, pronouns, adjectives, numerals, adverbs, conjunctions and postpositions, but the nouns and adjectives were not fully differentiated from one another and there were still some residual traces of an earlier stage of the language in which nouns/adjectives and verbs had not been fully differentiated. So far as creating new words by elongating existing words is concerned, Turkish, unlike some other languages, operated entirely by adding suffixes at the end of the word and had a very fully developed system of suffixes. These included, in addition to conjugational, declensional and possessive suffixes, suffixes for creating new verbs from existing verbs and from nouns/adjectives and for creating new nouns/adjectives from existing nouns/adjectives and from verbs, as well as some more specialized suffixes. This morphological system had been in existence for a long time, since in the earliest stage of the language which is known to us we find some suffixed words which carry suffixes which were no longer "productive," that is had ceased to be used currently to form new words, and some suffixed words which, when the suffix had been detached, were no longer words in current use.

The speakers of this language had been in contact, clearly for a long time, with communities speaking languages belonging to completely different families, and had borrowed words from them. Some of these loan words, and the languages from which they were borrowed, have

already been identified, and there are no doubt more to be found. Such early loan words are of great interest, since they throw some light on the history of the Turkish-speaking peoples at a date earlier than is otherwise accessible.

The further development of the Turkish languages can be sketched quite briefly, since it lies outside the main theme of this book. The position in the eleventh century is described by Kāşğarī in the Introduction to his *Dīwān*, and though the description is obviously incomplete, there is good reason for supposing that it is accurate as far as it goes. He was aware of the existence of a number of Turkish languages or dialects, most of them unwritten. In the north-east there was at least one language which had made the sound change initial **y-** > **c-**, possibly as I have suggested Ottuz Tatar. The other languages of the steppes were presumably similar to, perhaps even identical with, the late Uyğur of which texts have survived in Sinkiang. By this time Türkü and Uyğur-A seem to have disappeared, at any rate as written languages. But Kāşğarī distinguished between what he calls "Uyğur" and the official language of the Karakhanid dominions which he calls Xākānī. This suggests that the founders of the dynasty belonged to some tribe other than the Uyğur. It has been suggested that they were Karluk, but this cannot be taken as certain. By this time the Karakhanids were Moslems and Xākānī was written in the Arabic script. Kāşğarī has given us a very complete account of this language and we have one literary text of major importance in it, the *Kutadğu: Bilig*, of which more later. Kāşğarī says that there were important differences between this language and the Old Oğuz (in his terminology *al-Ğuzz*), the language of the Oğuz elements who had settled in the neighbourhood of the Aral Sea, and some of whom had, by the time he wrote, moved south and become the ruling classes under the Great Selcuk dynasty. Further west on the middle Volga at least one **l/r** language, Bulgar, to which he makes only occasional references, was well established, and in the steppes to the south Kıpçak, which was closer

to Old Oğuz than to any other language. Kāşğarī gives us some information about the phonology, grammar and vocabulary of Old Oğuz and Kıpçak and also much scantier information about some of the other languages, but the only continuous texts surviving from this period are the Uyğur and Xākānī texts mentioned above.

The Mongolian invasion naturally caused an enormous upheaval and widespread movements of tribes and whole peoples, the immediate effects of which it is impossible to estimate. One consequence of this upheaval was a considerable change in tribal nomenclature. For example it appears from the *Cāmi'u'l-Tawārīx** that the Turkish tribe with which the Mongols were most closely in contact at this time was the Tatar. In consequence this name became in the Mongolian period a generic term not only for Turks but also, and perhaps even more particularly, for Mongols. The name still survives on the middle Volga for the people in that area (the "Kazan Tatars") who speak a language quite different from Chuvash, the only surviving l/r language. This Tatar language is probably merely a later form of Kıpçak, but it and some, though not all, of the other descendants of Kıpçak have made the sound change initial y- > c-. It is not impossible that the name and this sound change commemorate the arrival in this area of a section of the original Tatars with the Mongolian armies, but it would be rash to base so important a conclusion on such slender evidence.

So far as written texts which can be dated to the century or two following the Mongolian invasion are concerned the position is broadly as follows:—In Sinkiang late Uyğur continued to be written in the old Uyğur script and we have legal and commercial documents at any rate as late as the fourteenth century and Buddhist religious texts even later (see Chapter V (1) (iv)).

In the centre there was a slow evolution from the Xākānī of the eleventh century to the so-called Çağatay of

* See *Rashid-ad-din, Sbornik Letopisey*, 3 volumes, Moscow-Leningrad, 1952-60, Vol. I, pp. 101 foll. I am indebted to Dr. J. A. Boyle of Manchester University for this reference.

the fifteenth century, and several texts from between these two dates have survived. These were all originally written in Arabic script, but one or two transcriptions in the Mongolian Official Alphabet (see Chapter IX) have survived. These texts include the Verse and Prose Prefaces of the *Kutadğu: Bilig*, the *Atabatu'l-Haqā'iq**; the *Nahcu'l-Farādīs*†, and the *Qiṣaṣu'l-Awliyā* of Rabğūzī, of which a good many editions, none of them critical, have been published.

There was during this period a cultural and literary centre in Xwarazm to the south and south-east of the Aral Sea, where an Oğuz language usually called Türkmen was written in the Arabic script. The most important literary specimens of this language are Quṭb's version of the *Xusrāw wa Şirīn* of Nizami‡ and Xwarazmi's *Muḥabbat Nāma*§, both written in about the middle of the fourteenth century. It seems likely that the enigmatic *Oğuz Name*, of which a transcription in the Mongolian Official Alphabet survives (see Chapter IX), and from which a sentence is quoted in the *Sanglax*, is also in this language.

Four Arabic vocabularies of the Turkish languages spoken by the slaves (*mamlūks*) imported into Egypt from southern Russia have survived, ranging in date from A.D. 1245 to the fifteenth century. Most of the entries are in Kıpçak, but some Türkmen material is included. The only continuous Kıpçak texts of the fourteenth century are probably one or two Christian hymns, sermons, *etc.* in the *Codex Cumanicus*, a handbook of the Kıpçak language called Koman compiled by Christian missionaries in south Russia in the early fourteenth century. There are some rather later Kıpçak texts in Armenian script.

* Edited by R. R. Arat, *Atebetü'l-hakayık*, Istanbul, 1951.

† A facsimile of the best manuscript with a preface by J. Eckmann has been published, *Nehcü'l-feradis*, Ankara, 1956.

‡ Published by A. Zajączkowski with facsimile and vocabulary, Warsaw, 1958-61.

§ Published by T. Gandjei with facsimile, translation and vocabulary, Annali dell'Instituto Universitario Orientale di Napoli, Vols. VI-VIII, 1958-59; by A. M. Shcherbak, Moscow, 1959 (one MS. only with translation), and by E. N. Nadzhip, Moscow, 1961 (text, translation, vocabulary and facsimile).

In the thirteenth century the Oğuz language spoken in Anatolia began to be written in Arabic script. The very early texts are not entirely homogeneous, and there were probably one or two Oğuz dialects spoken in this area at that time. The earliest specimens are a few Turkish verses scattered through the voluminous Persian works of the great mystic poet Mawlānā Calālu'l-Dīn al-Rūmī, who was born in Balkh in A.H. 604 (A.D. 1207-8) but fled to Anatolia as a young man to escape the Mongols and died in Konya in 672 (1273-4). It has been the custom, and is formally correct, to call the language written by him and his son Sulṭān Walad Bahā'u'l-dīn Aḥmad "Selcuk," since at that time the Selcuks of Rūm were still reigning at Konya. Similarly it would be formally correct to call the language in which Qāḍī Burhānu'l-dīn of Sivas, who died in A.H. 800 (A.D. 1395-6), wrote his poems "Türkmen"; but as there is a continuous literary tradition in Anatolia from the thirteenth century onwards, and as Osman ('Uṯmān) I became an independent monarch, though not the ruler of more than a small part of Anatolia, in A.D. 1299, it is more convenient to call all the earlier stages of this language "Old Osmanli," irrespective of minor dialect variations.

This is not of course a complete list even of the literary texts, let alone the religious texts and administrative and other documents, in Turkish languages from the period between the twelfth and fourteenth centuries inclusive, but it is a broad indication of the material which is available for tracing the history of those languages during this period.

If it be accepted that direct links exist between Çağatay and modern Uzbek, between Kıpçak and the various languages written during the intermediate period in southern Russia (Kazan Tatar, Krim Tatar and Karaim) and between Türkmen and Azeri, and that the Türki of Sinkiang, which was written in Arabic script at any rate as early as the eighteenth century and no doubt earlier, is directly descended from Xākānī, although the intermediate links are missing, then so far as I am aware no other Turkish language

significantly different from those mentioned above became a written language until the nineteenth century.

It was only in the nineteenth or even twentieth century that the remaining Turkish languages which exist to-day and are markedly different from those mentioned above, that is the north-eastern group of languages in southern Siberia and Tuva, and Kazax and Kırğız in the north centre, became written languages. The exact line of descent of these languages has still to be worked out; some progress has been made but in the absence of any written texts prior to the nineteenth century the process is a difficult one.

It is in the nature of things impossible by looking at a language at any single stage of its development to judge which native words in it have existed for a long time and which were invented only recently. This can be done only by comparing different stages of the language and then only tentatively. The texts which have survived from the period anterior to the eleventh century are rather inconsiderable in bulk, and the longest of them, translations of Buddhist scriptures, have a rather restricted vocabulary. There are, however, two eleventh century texts, Maḥmūd al-Kāşğarī's *Dīwān luğāti'l-Turk*, the earliest Turkish dictionary, and Yusuf Xaṣṣ Ḥācib's great didactic poem of over 6,000 couplets, the *Kutadğu: Bilig** which between them gives us a very extensive vocabulary of the most important language of that century, for which Kāşğarī's name "Xākānī" has been generally accepted. A careful analysis of the vocabulary of the texts anterior to the eleventh century shows that a few genuine Turkish words used in them seem to have fallen out of use, or at any rate are not recorded as still in use in Xākānī. On the other hand it is impossible to say with any confidence that any words which are not recorded earlier than in Xākānī were invented in

* I shall make frequent references to these two works. The first will be cited as "Kāşğarī," precise references like *I*, 28 being to B. Atalay's translation *Divanü Lûgat-it-Türk Tercümesi*, 3 volumes, Ankara, 1940-41. The second will be cited as "K.B." followed by the number of the couplet in R. R. Arat's critical edition, İstanbul, 1947.

the eleventh century; they may well have existed before that date but not been used in the texts which have come down to us.

The texts in Runic script (see Chapter V (1) (1)) contain very few loan words, not many more than those mentioned in Chapter I, but the other pre-eleventh century texts, Buddhist, Christian, Manichaean, medical, legal, commercial and miscellaneous, are almost all, except the last three categories, translations from other languages and contain many loan words borrowed directly or indirectly from Indian or Iranian languages. Very few of these loan words survived in Xākānī or later languages; they were nearly all religious technical terms, and almost all the Turkish texts which have come down to us from the eleventh century onwards to the recent past were written by Moslems and used the international Moslem religious vocabulary.

Although, for the reasons given above, it is impossible to identify any newly invented native words in Xākānī, there are a good many Xākānī words which seem to have become obsolete soon after the eleventh century and cannot be traced at any subsequent date. It is, of course, possible that some of them survived for a time, or indeed still survive, but happen not to have occurred in later texts.

It is in Xākānī that we find the first representatives of the great mass of Arabic and Persian loan words which entered all the Turkish languages except perhaps those in the north east where Islam never penetrated. Most of them are found in the *K.B.* and very few in Kāşğarī, but Kāşğarī, whose dictionary contains words from several other dialects besides Xākānī, mentions and identifies some everyday Persian words which had found their way into Old Oğuz, for example **ören** " ruined, deserted," which is a corruption of Persian *wayrān*. A few Persian and Arabic loan words and proper names also occur in the late Uyğur legal and commercial documents.

In the thirteenth century a completely new class of loan

words began to appear in various Turkish languages, more in some than in others. The invasion of central and western Asia by the Mongols was followed by the invasion of the Turkish languages by Mongolian loan words. The earliest of these are found in the late Uyğur legal and commercial documents, in the Arabic vocabularies of Kıpçak and Türkmen, in the *Codex Cumanicus* and in the so-called *Oğuz Name* already referred to, but the main bulk of them first appears in Çağatay. I devoted a good deal of attention to them in the Introduction to the *Sanglax*, and included a list which is probably almost complete of the Mongolian loan words found in that work.

A comparison between the surviving vocabulary of Xākānī and the vocabularies of the post-eleventh century texts mentioned above and of Çağatay shows not only that some Xākānī words seem to have fallen out of use, but also, even after the Arabic, Persian and Mongolian loan words have been segregated, that there are a good many new words which are obviously good Turkish but cannot be traced at an earlier date. Some of them may well have existed in Xākānī, but happen not to have been used in any text which has survived from the eleventh century or earlier. It occasionally occurs that such a word can be proved to have existed at a much earlier date because it occurs as an early loan word in Mongolian or Magyar or both. A case in point is **köşek** " a camel colt " which does not seem to be traceable before it appears in Old Osmanli in the fourteenth century. It appears however as *gölige/gölöge* " a puppy " in Mongolian and *kölyök* " a kid, or young animal of another kind," in Magyar, and as these forms show that the word must in both cases have been borrowed from an **l/r** language, **köşek** meaning probably " a young animal " must have been a Turkish word before the split between standard and **l/r** Turkish. One or two of the words attested for, say, Çağatay but no earlier may not be genuine words at all, but merely the result of scribal errors. But the rest of these untraceable words must have been invented after the eleventh century. In Chapter VI I shall mention some

criteria which help to determine whether such words are likely to be genuinely Turkish.

I can perhaps usefully conclude this chapter by suggesting what the future course of Turkish lexicography should be. The first stage, that on which I am at present engaged, is to compile a list of all Turkish words known to have existed in or before the eleventh century, noting the earliest occurrences and meaning of each word and in appropriate cases sketching its subsequent history, but excluding all loan words except those very old ones which can throw light on the external contacts of the Turkish peoples in the prehistoric and earliest historic periods. This is an essential spring-board for any future scientific work on the language.

The next stage will be to carry the collection of vocabulary material forward to, say, the fifteenth century, which involves adding to the eleventh century material any genuine Turkish words which appear for the first time in the texts mentioned above or in Çağatay or Old Osmanli, and in appropriate cases recording further developments in the meanings of the older words where this has not already been done.

There is of course a great mass of material dating from the sixteenth to eighteenth centuries inclusive, but it is all in languages the vocabularies of which will have been compiled in the second stage of the work or in languages closely related to them. It is difficult to see the utility of compiling a comprehensive dictionary of all the Turkish languages as they existed at any later date than the fifteenth century.

When Radloff compiled his great *Opyt* there were in existence only a few dictionaries of various Turkish literary languages and practically none of the contemporary unwritten languages. He had access, except when compiling the first volume, to the Runic inscriptions, but none of the manuscripts in Türkü, Uyğur or Uyğur-A had been discovered, and he did not have access either to Kāşğarī or to the *Sanglax*. His courageous attempt to collect in a single volume all the vocabulary material that he could lay his

hands on has been, and still is, of untold value to subsequent generations of students. But the time for work of this kind is now past.

The modern Turkish languages all have characters of their own. There are phonetic differences and differences of spelling between them which are often large, as well as differences of accidence, syntax and vocabulary. Each of them has its own repertoire of loan words dependent on the cultural and other contacts which its speakers have had with peoples speaking other languages. If such a language has a long literary tradition behind it, like Osmanli/Republican Turkish or Çağatay/Uzbek, a historical dictionary showing the developments of meaning of the native words, with references back to the basic vocabularies of the fifteenth and earlier centuries, and the dates and circumstances in which various loan words entered the language will be of great interest to the speakers and students of that language. But it is hard to see the utility of including all the material in all these historical dictionaries in one vast corpus, which would be comparable to, say, a combined historical dictionary of all the Indo-European languages.

On the other hand historical dictionaries of individual Turkish languages might well provide material which would serve to amplify the basic dictionary of eleventh century Turkish. I have just pointed out that a comparison between Old Osmanli and the early loan words in Mongolian and Magyar proves that **köşek** must have been a Turkish word nearly two thousand years ago. A comparison of historical dictionaries between themselves and with lists of early Turkish loan words in other languages might make it possible to add other words to the basic dictionary. As I pointed out in the Introduction, in Turkish more perhaps even than in other languages words have a habit of, so to speak, going underground and reappearing centuries later in some modern spoken language, although they have ceased to be used in literary texts.

CHAPTER III

THE TRANSCRIPTION OF TURKISH LANGUAGES

As I understand the matter, transliteration and transcription are entirely different processes. Transliteration is the substitution of the letters or signs of one alphabet for those of another, in the case of Turkish usually the substitution of letters of the Latin or Cyrillic alphabets, with the addition of modified letters, for the letters and vowel signs of one of the alphabets enumerated in Chapter V. The purpose of transliteration may be no more than to make it possible to print the whole of an article in congruous type founts and so avoid the expense of printing with mixed founts, without depriving the reader of the possibility of knowing exactly the form of the original text in the other alphabet. Alternatively it may be, so to speak, to reduce to a common denominator a set of extracts from texts written in different alphabets. In almost every alphabet, certainly in all the alphabets enumerated in Chapter V, some of the letters used are really polyphonic, that is represent more than one sound. And so even before he begins his work the transliterator must formulate a whole series of stated or unstated conventions, some of which are, phonetically speaking, misleading. Thus he may, for example, decide to use Latin **b** to represent the Arabic letter *bā*, which would probably be correct if he were transcribing an Arabic text but would quite often represent the wrong sound if he were transcribing Turkish or Persian. Thus the transliterator must model himself on the legendary Hong Kong tailor who, given a roll of cloth and an old suit to copy, produced so faithful a copy that it reproduced all the tears, stains and abrasions of the original.

Transcription, on the other hand, is an attempt to represent in one alphabet the actual sounds which the writer of a text intended to represent in another. The

essential difference between it and transliteration is that the latter proceeds on the principle of substituting one letter, and always the same letter, of one alphabet for one letter of the other, while the former is an attempt to interpret those signs and letters of the other alphabet which are polyphonic and to substitute for each sign or letter of the original the letter of the transcription alphabet which is most appropriate in each case, without being too much worried by the fact that in certain circumstances one letter in the transcription alphabet may represent different letters in the original alphabet, or even represent a sound, usually a vowel sound, which is not represented at all in the original text.

The human voice is capable of producing such a wide range of sounds (and in tonic languages tones) that it is almost impossible to devise an apparatus of printed letters capable of representing them faithfully on paper. And so at best any transcription is more or less imperfect, even when the language concerned is a modern one with which the transcriber is perfectly familiar, and even when he employs some kind of phonetic alphabet so complicated that it is difficult and costly to set up and almost impossible to read intelligently. Transcription becomes a proportionately less and less accurate medium for representing the original sounds, both as the range of letters in the transcription alphabet is narrowed, and as the transcriber is less and less familiar with the actual sounds of the text which he is transcribing, either because he is imperfectly acquainted with the language or because the text is in a dead language the phonetics of which cannot be confidently reconstructed. At its worst a transcription may be so imperfect a representation of the original text as to leave the reader in great doubt what sounds the transcriber really intended to represent.

Nevertheless even an indifferent transcription is more useful than a perfect transliteration. Transliteration merely passes the buck to the reader and leaves him to grapple with a whole range of problems, for which it is really the duty

of the writer to propose solutions. Provided that a transcriber presents his solutions to these problems with becoming modesty and states his reasons for them, he cannot mislead the reader and may even inspire him to produce better ones. And so I make no apology for proposing a system of transcription for the Turkish languages, knowing from the start that it must be a very imperfect one, and that for two reasons. The first is that our knowledge of the phonetics, particularly of early Turkish, is so imperfect that it would be foolish to use anything more scientific than a very simple transcription alphabet, sufficiently refined to ensure that each letter represents a sound or sounds distinct from those represented by any other letter, but not so refined as to provide separate representation for sounds so close to one another that there is really no means for determining which of them should be used in any particular case. We can be reasonably sure that there were in pre-eighth century Turkish more than nine vowel sounds (quite apart from differences of length) and more than twenty-three consonantal sounds, but that is about as many as we can distinguish with any confidence. The second is that, at any rate so far as earlier stages of the language are concerned, the material which is at our disposal is for the most part so ambiguous as to permit a wide margin of differences of interpretation.

When European scholars first began to study the Turkish languages, no such language was written by its speakers in a Latin alphabet, and almost every student devised an alphabet of his own based on that alphabet. The earliest, and least scientific, of these alphabets was that used by the Italian monks who wrote the main part of the *Codex Cumanicus* early in the fourteenth century. The alphabet with some additional special letters which was used by the German monks who made additions to that manuscript some years later was more satisfactory, but far from perfect. Since then several conventional alphabets of this kind have established themselves and are currently in use, a good deal of variety existing particularly in the representation of

both **e-** sounds, back **ı**, the velars, fricatives and affricates. A new situation was created thirty years ago when the Turkish Republic adopted an Official Alphabet based on the Latin alphabet. It seems reasonable to suggest in the interests of uniformity which is so important in international studies, that all students of the Turkish languages, other than those in the Soviet Union who will no doubt find it more convenient to use an alphabet based on the Cyrillic alphabet with some additional letters of the type used in writing Turkish languages spoken in the Soviet Union, should abandon the various conventional alphabets now in use and agree to use the Turkish Official Alphabet with such additional letters as may be required either to represent sounds which do not exist in Republican Turkish, or to differentiate between two sounds which are represented by the same letter in that alphabet.

The alphabet which is set out below is constructed on those principles. In this chapter I do not wish to do more than to set out that alphabet and explain what sounds the various letters are intended to represent. The question whether any particular sound actually existed in a particular Turkish language at a given date will be discussed later.

VOWELS	Back,	Short **a, ı, o, u.**
		Long **a:, ı:, o:, u:.**
	Front,	Short **e, é, i, ö, ü.**
		Long **e:, é:, i:, ö:, ü:.**

These letters are a good deal more conventional, and possibly polyphonic, than the consonants listed below. In the first place, while the existence of long vowels in early Turkish and in some modern Turkish languages cannot be disputed, the effect, if any, which length had on the quality of the vowel is uncertain. In the second place, there is great uncertainty even about the quality of the short vowels in early Turkish, and there may well have been different qualities of some of them in different languages.

A was presumably a low back sound, something like the first vowel in English *father*, or German *vater*. It is not

possible to say whether a difference of length involved a difference of quality.

I was presumably a back sound like the **ı** in Republican Turkish, for which there is no real equivalent in any Western European language. It is not possible to say whether a difference in length involved a difference in quality.

Considerable doubt exists regarding the quality of the **o** in early Turkish. In Republican Turkish **o** is an open back sound like the vowel in English *not* or German *Gott* and that may have been the sound in early Turkish. Alternatively it may have been a more closed sound like that in English *hope*, that is the sound represented in Republican Turkish by **oğ** in words like **oğlu.** There is of course a possibility that it was more open when short and more closed when long.

U was probably a closed back sound like the vowel in English *boot* or German *gut*, that is the sound of Republican Turkish **u.** It is less likely to have been a more open sound like the vowel in English *foot*, but it may have been more open when short than it was when long.

E was presumably, like Republican Turkish **e,** an ordinary open *e* like the vowel in English *get* or the first vowel in German *hätte*, but it is difficult to guess whether length made any difference to the quality of the sound.

É was presumably a closed *é*, like the vowel in English *gate* or, better still, German *geht*; here again it is difficult to guess whether length made any difference to the quality of the sound.

The quality of **i** is most uncertain. It may, like the Republican Turkish **i,** have been an open sound like the vowel in English *bit* or German *bitte*; alternatively it may have been a more closed sound like the vowel in French *vite* or Italian *vita*. Here again a difference of length may have involved a difference of quality.

Ö and **ü** probably had the same sounds as those represented by the same letters in Republican Turkish and German, which do not exist in standard English. Here

again a difference of length may have involved a difference of quality.

CONSONANTS

	Plosive		Fricative		Nasal	Affricate		Sibilant		Semi-vowel
	v.*	u.	v.	u.		v.	u.	v.	u.	
Labial	b	p	v	f	m	—	—	—	—	w
Dental	d	t	d̲	—	n	—	—	z	s	—
Denti-palatal	—	—	—	—	—	c	ç	—	—	—
Palatal	—	—	—	—	ñ	—	—	j	ş	y
Post-palatal	g	k	—	—	} ŋ	—	—	—	—	—
Velar	—	k	ğ	x		—	—	—	—	—

Liquids:—l, r. Aspirate h.

* v. means "voiced"; u. means "unvoiced (or voiceless)."

This table will, I hope, be self-explanatory. There is perhaps one defect in the system of transcription proposed. Following the precedent set by the Turkish Official Alphabet I use **k** for both the post-palatal and the velar unvoiced plosive; this cannot lead to any confusion since the quality of the consonant is determined by the adjacent vowel. I am not entirely sure that I have found the right solution for the corresponding voiced sounds, plosive and fricative. The balance of evidence is that the voiced velar sound in early Turkish, as in Arabic, was a fricative which I represent by **ğ**, but it is possible that this was sometimes a plosive. For the equivalent post-palatal sound I have used **g**; this seems normally to have been a plosive, but it may sometimes have been a fricative. The existence of an unvoiced velar fricative, which I represent by **x**, in early Turkish is uncertain, but there does not seem to be any evidence for the existence of an unvoiced post-palatal fricative.

It is unlikely that there were in early Turkish half-voiced plosives, as there are in Chuvash (see J. Benzing, *Das Tchuwaschische* in *Philologiae Turcicae Fundamenta*, Wiesbaden, 1959), and I have made no provision for their representation.

Some additional letters are required to transcribe Arabic loan words in Turkish. These are the unvoiced dental

fricative, **ṯ**, the " emphatic " sounds, **ḍ**, **ṭ**, **ḥ**, **ṣ** and **ẓ**, the glottal stop ', and the '*ayn* '. It is also necessary, in transcribing Arabic, to distinguish between the unvoiced velar plosive, **q**, and the unvoiced post-palatal plosive **k**. In transcribing Arabic and Persian, as opposed to Turkish, I mark the long vowels by a superscribed line **ā**, **ē** (which is, strictly speaking, incorrect as the Persian *e* is a closed *é*), **i**, **ō** and **ū**, since this is the normal method of representing these sounds. The use of a colon (:) to indicate length in Turkish vowels is dictated by the inconvenience and unsightliness of a long sign above **é**, **ö** and **ü**.

CHAPTER IV

THE ARRANGEMENT OF WORDS IN A HISTORICAL DICTIONARY OF TURKISH

The purpose of a dictionary is to enable the user to find any word and its meaning with a minimum of difficulty in a predetermined place. Where a language has an established system of spelling, as almost all modern languages have, anyone who knows how to spell and is not conducting his search on purely phonetic principles is sure of finding the word he is seeking at the first attempt. He will not of course do so if, for example, he assumes that English is spelt on phonetic principles and looks for " honour " under O or " use " under Y.

But when a dictionary is compiled on diachronic principles, that is sets out in a single list the forms which the same word had at different times, or is intended for use with texts written in an alphabet or alphabets other than that used in the dictionary, a new situation arises. The dictionary must in effect be arranged on phonetic principles and must cater for people who want to find in it, for example, a word written in unvocalized Arabic script, for the transcription of which there is in theory a wide range of alternatives. In so far as it is possible, the principles followed must be sufficiently elastic to make it possible to group under a single heading words which at different times, or even at the same time in different dialects or languages, contained slightly different sounds, and a predetermined place for each such heading must be laid down in advance. In the last resort when there are wide differences of spelling, for example when a word begins with, say, **b-** in one language and **m-** in another, it will be necessary to supplement the main heading with one or two cross-references, but it is always frustrating to look for a word in one place only to be told to look for it in another, and

the number of cross-references should be reduced to a minimum.

The problem was one which confronted me at the very beginning of my lexicographic work. At that time my plans were at their most comprehensive, and my purpose was to include under a single heading all the forms which a word had assumed from the eighth to the twentieth centuries, that for " good," say, starting with **edgu:** and finishing with forms like **iyi.** It was obvious that if every form of every word was to be listed in strict alphabetical order the cross-references would take up more space than the main headings. Some form of grouping was obviously essential, and after a great deal of trial and error I finally evolved a system which seems to me to be the most practical one that can be devised. The task was of course enormously simplified when I decided to exclude, except in brief summaries under some words, all forms later than the eleventh century. For example, instead of having half a dozen cross-references to **edgü:** it became unnecessary to have any, since the first divergent form, **eygü,** is not noted before the fourteenth century. As against this, it has become clear to me that an almost essential part of any dictionary of early Turkish will be a reverse index of meanings. The work would be much more valuable if, for example, someone who wished to discover the antecedents of **iyi** " good " in Republican Turkish could get on to the earliest form **edgü:** simply by looking up " good " in the reverse index. This index would also have a negative value in the sense that the absence of a word in it would indicate that there was no early Turkish equivalent for that word. For example, anyone investigating the twentieth century Anatolian word **itavun** who looked up " partridge " in the index and found no word even remotely resembling **itavun** under it would realize that that word was probably not really Turkish and would be well on the way to discovering that it is really a Mongolian loan word.

The system which I finally evolved was curiously like that used in an Arabic dictionary, though the resemblance

is really fortuitous. I started by arranging the letters of the alphabet in a number of groups:—the vowels; the labials (**b, p, v,** but not **m** which was classified as a nasal); the affricates (**c** and **ç**); the dentals (**d, d̲** and **t,** but not **n**); the velars (**ğ** and **k**); the post-palatals (**g** and **k**); **l; m;** the other nasals (**n, ŋ** and **ñ**); **r; s; ş; y; z.** In order to provide running headings for the dictionary, I converted these groups into code letters:—A, B, C, D, Ğ, G, L, M, N, R, S, Ş, Y, Z, and a series of sections was devised, each headed by one or more of these code letters, so that by converting any given word into a series of code letters its place in the dictionary could be found. I laid it down that in doing this account should be taken, so to speak, only of the bone structure of the word, that is the initial vowel, if any, the consonants, and the final vowel, if any, other vowels being disregarded. I then decided that, in order to reduce the number of sections to manageable proportions, no section should be headed with more than three code letters, and that separate sections should be provided for monosyllables (Mon.), dissyllables (Diss.) and longer words (Tris.) and within these groups separate sections for words other than verbs and for verbs, the latter marked V. Thus the first few main sections of my dictionary are Mon. A (containing words consisting of a single vowel like **u:** " sleep "), Mon. V.A- (containing verbs consisting of a single vowel like **u:-** " to be able "), Mon. AB (containing words consisting of a vowel followed by a labial like **a:v** " wild game "), Mon. V.AB- (containing verbs like **év-** " to hurry "), Diss. ABA (containing words like **apa:** " ancestor ") and so on. By this system the whole vocabulary is broken down into a number of sections (theoretically nearly 3,000, in practice very much less) and only a very few sections contain as many as a dozen words. Diss. V.ADL- is quite exceptional in containing as many as forty, **adal-, ad̲ıl-, atıl-, édil-, ıd̲ıl-, itil-** and so on. Thus the number of words in the section which must contain a word, if it existed in early Turkish, is comparatively small, and the means for determining in which section a word is

to be found are very simple. For example **atlığ** and **otlak** will both be found under Diss. ADL, **katığlan-** under Tris. V.ĞDĞ- and so on. Even so it was necessary to lay down principles for determining the order of words within each section. Obviously the first determinant was any sound in a word beyond those contained in the heading; for example in the section Diss. V.ADL-, **adla:-** (ADL(A)-) must come later than **ütül-** (ADL-) and earlier than **atlat-** (ADL(D)-). More difficult questions arose when several words could be converted into the same code letters, for example **adal-, adıl-, atıl-, édil-, ıdıl-** and so on, all ADL-. Here it seemed to me that the main factor to be taken into account was the probable differences of spelling in the original texts, and the purpose was to arrange the words in such a way that someone who knew the general pronunciation of a word but not its exact pronunciation would look in the right place first. In view of the wide discrepancies between the efficiencies of the various alphabets concerned a compromise was inevitable. Thus for example the Uyğur script always, and the Arabic sometimes, have different scriptions for initial **a-; e-; é-/ı-/i-; o-/u-** and **ö-/ü-,** but the Uyğur and Brāhmī scripts are almost alone in distinguishing between **a-** and **e-,** and Brāhmi in distinguishing between **o-, u-, ö-** and **ü-,** and Uyğur almost alone in not distinguishing between **-d-** and **-t-.** On the other hand Runic is almost the only script which distinguishes between **-b-** and **-p-.** These spelling idiosyncrasies will be explained in detail in the next chapter. Thus, after reviewing the whole situation, it seemed to me that the most sensible thing so far as consonants are concerned would be to treat **b** and **p** as identical, but to arrange other consonants in a single group in their natural alphabetical order, for example the dentals in the order, **d, d, t.** So far as vowels are concerned it seemed to me that they should be arranged in the order **a, e, é, i, i, o/u** (which frequently interchange), **ö/ü** and that the first vowel in a word, whether initial or not, should usually be regarded as more important that a subsequent consonant for determining the

order of words. Thus I have placed **at-** before **éd-**, but **üḏ-** before **öt-**, because it is seldom possible to distinguish between **ö** and **ü** in the alphabets used to write Turkish, but usually easy to distinguish between **ḏ** and **t**. Similar empirical rules were devised for dealing with other doubtful cases. It is very likely that other people using these general principles would in some cases put a few words in a different order, but in practice the differences would be so slight that no difficulty would be experienced in finding a particular word, provided of course that it was in the dictionary.

This system is usually not applicable to a dictionary of any language which has a generally accepted system of spelling and is written in the same alphabet as the dictionary itself, but it has great advantages for any dictionary in Latin letters of a language or dialect habitually written in some other form of script except perhaps Cyrillic.

Years of experience have shown me that great frustration is caused by adherence to a strict alphabetical order, whether it is a normal one or a highly idiosyncratic one like that followed in Radloff's *Opyt*, when it results in, say, the same Çağatay word being entered in two or more quite different places in different transcriptions; a case in point is **çöşkür-** and **çüşkür-**, which are in fact the same word, in the *Opyt*; and it is a great comfort, if there is genuine doubt whether a word in, say, the *Kutaḏğu: Bilig* is spelt with an **-o-** or a **-u-**, to know that in either event it will be found in the same place in the dictionary.

CHAPTER V

THE EVIDENCE REGARDING THE PHONETIC STRUCTURE OF PRE-EIGHTH CENTURY TURKISH

The raw material which is available to help us to reconstruct the phonetic structure of pre-eighth century Turkish falls into four classes:—

(1) Early Turkish texts.

(2) Turkish names, words and phrases in foreign authorities, mainly Chinese and Byzantine.

(3) Turkish loan words in foreign languages, in particular Mongolian and Magyar.

(4) Modern Turkish languages.

(1) and (2), and to some extent (3), share the common disadvantage that they were written down a long time ago in scripts made up of letters, signs or characters the exact phonetic value of which can hardly be more than a matter of conjecture; (4), and (3) to the extent that modern forms of the languages are used as evidence, share the common disadvantage that there is a large chronological hiatus between the eighth century and the present day and that it is too much to hope that no phonetic changes have occurred during this long period.

(1) EARLY TURKISH TEXTS

These have come down to us written in six (or, if Sogdian and Uyğur are taken as different, seven) different scripts each with its own merits and defects. They are (i) "Runic"; (ii) Brāhmī; (iii) Tibetan; (iv) Sogdian and Uyğur; (v) Manichaean Syriac, and (vi) Arabic.

(i) THE RUNIC TEXTS

When the first monuments in Runic script were rediscovered in the eighteenth century it was for a time

believed that they were written in Scandinavian runes. It was soon realized that this was an error, but the name has stuck and is in fact more convenient than any other which could be devised.

Apart from a few short inscriptions on jugs, spindle-whorls and the like, which are philologically unimportant, the texts in Runic script fall into four classes:—*

(a) Funerary monuments in Outer Mongolia erected over the graves of Türkü and Uyğur rulers and high officers in the eighth and ninth centuries. They can conveniently be called collectively " the Orkhon inscriptions," although in fact only some of them are in the basin of that river. Except for three monuments, the inscriptions at Şine Usu, Karabalğasun (fragmentary) and Sucı, which are presumably written in Uyğur, the language of all these inscriptions is Türkü. These texts were prepared for the masons by educated people and are written in good calligraphic scripts with very few obvious errors.

* The most comprehensive edition of these texts is H. N. Orkun, *Eski Türk Yazıtları*, 3 volumes and index, İstanbul, 1936-41 (cited as Orkun, *op. cit.*). This has since been to a large extent superseded by better editions of individual texts. A. von Gabain re-edited the inscription of Kül Tégin (cited as *KT*) in the Christomathy to her *Alttürkische Grammatik*, Leipzig, 1941 (cited as von Gabain, *ATG*), and S. Ye. Malov produced another edition of this inscription in *Pamyatniki Drevnetyurkskoy Pis'mennosti*, Moscow, 1951 (cited as *Malov*, 1951). I re-edited the the Ongin inscription in *The Ongin Inscription*, J.R.A.S., 1957. P. Aalto re-edited the inscription of Toñukuk (cited as *Toñ.*) in *Materialen zu den alttürkischen Inschriften der Mongolei*, Journal de la Société Finno-Ougrienne, LX, Helsinki, 1958. S. Ye Malov in *Pamyatniki Drevnetyurkskoy Pis'mennosti Mongolii i Kirgizii*, Moscow, 1959 (cited as *Malov*, 1959), re-edited the inscriptions of Bilge Kağan (cited as *BK*) and Küli Çor (cited as *KÇ*), the Uyğur inscription at Şine-usu (cited as *Şu.*), a number of shorter inscriptions from Mongolia, the five inscriptions from the Talas and a few other texts. One or two of the last are re-editions of texts previously published in his *Yeniseyskaya Pis'mennost' Tyurkov*, Moscow, 1952 (cited as *Malov*, 1952) which also contains better editions of all the Yenisei inscriptions in Orkun *op. cit.*, some additional ones, and a new edition of the Sucı inscription, the last without taking into account the corrections made by K. Grønbech in " *The Turkish system of kinship*," Studia Orientalia Ioanni Pedersen . . . Dicata, Copenhagen, 1953. *Malov*, 1951, also contains a new edition of the *Irk Bitig* (cited as *IB*), and I have proposed further amendments to the text and translation in *Notes on the " Irk Bitig "*.

It is very doubtful whether any of these editions can be regarded as absolutely final; there is probably not much left to be done with the Orkhon inscriptions or the manuscripts, but it is clear that the present editions of the Yenisei inscriptions are still most unsatisfactory and very little reliance can be placed upon them. A. M. Shcherbak, a Russian scholar who has specialized in this field, has recently been touring in Tuva, putting some inscriptions in a central museum and re-examining others. We may therefore confidently expect better editions of some of them in the near future.

(b) Monuments, mainly funerary, erected by tribal chiefs and the like and commonly called "the Yenisei inscriptions" because most of them have been found in the basin of the upper Yenisei and its tributaries. These texts were prepared by much less educated authors than the Orkhon inscriptions and some of them seem to be quite illiterate or frankly unintelligible. The masons who carved them were less skilled and the letters are clumsily formed and unsightly. They fall into two different groups separated by differences of spelling and orthography. The larger group, comprising Nos. 1 to 24, 40 to 46 and 49 to 51 in *Malov* 1952, all found in Tuva, and No. 25, found in Khakasia but bearing a Tuvan *tamğa*, are distinguished from the rest by the use of a special letter for closed **é** in words like **él** "realm," and by the retention of the original initial **b-** in words like **ben** "I." These inscriptions are commonly, and no doubt correctly, believed to have been erected by the Kırğız; they can conveniently be called "the Tuvan inscriptions," and the language in which they are written as "Old Kırğız." The second, and smaller, group comprises Nos. 26 to 39 and 48 in *Malov* 1952 and follows the same orthography as the Orkhon inscriptions spelling **él** either *il* or *l* and assimilating initial **b-** before a nasal as in **men** (for **ben**) "I." All these inscriptions come from Khakasia and can conveniently be called "the Khakasian inscriptions." The authors were presumably not Kırğız, and the fact that No. 37 refers to the Türgeş tribe and No. 48 commemorates a man called **Tölis Bilge:** suggests that they may have been erected by descendants of Western Türkü tribesmen who took refuge in the Altai when the Karluk "destroyed" the Türgeş in A.D. 766. Owing to the presence of the special letter for **é** in the Tuvan inscriptions and their generally primitive appearance it was at first supposed that they must be older than the Orkhon inscriptions, and dates as early as the sixth, and even the fifth, century were suggested for them. For some time past a few scholars have been suggesting that these dates are not consistent with the contents of the inscriptions, and

that roughness and unsightliness are more often an indication of provincialism or decadence than of great age. The question has now been settled once and for all by L. R. Kyzlasov, who has examined a number of these inscriptions in their archaeological context. In a brilliant article, *Novaya datirovka pamyatnikov Yeniseyskoy pis'mennosti*, Sovetskaya Arkheologiya, 1960, part 3, he has given reasons for supposing that, with two exceptions, these inscriptions cannot be earlier than the ninth or tenth centuries and that some may be even later. He dates *Malov* 1952, No. 40, to the eighth century for archaeological reasons, which are no doubt adequate, but the text is very short and incoherent, and seems to have been mistranscribed. There is a definite possibility that the stone was re-used at a later date than the objects with which it is associated. He dates *Malov* 1952, No. 32, to the end of the seventh century, but this dating rests on two false hypotheses. The first is that both this inscription and the Ongin inscription mention a man called Taçam; the second is that the Ongin inscription was the memorial of Elteriş Kağan who died in A.D. 691. I have, however, given reasons in *The Ongin Inscription* for supposing that it was erected between A.D. 732 and 734 and does not contain the proper name Taçam. Moreover, even if the name did occur in both inscriptions, they need not necessarily refer to the same man. The edition of No. 32 seems to bear very little relation to the text shown on the squeezes published in Radloff's *Atlas der Alterthümer der Mongolei*, Pt. III, St. Petersburg, 1896, plates 93 and 94. From the squeezes, which are extremely bad, it looks as if the stone had originally been erected as a funerary monument with one inscription carved on it and had later been re-used for a second inscription in letters of a different size. The two seem to have become hopelessly confused in the edition and really make no sense at all. It is possible that expert examination of the stone itself might produce some kind of a coherent pair of texts. There does not seem to be any good reason for supposing that either is earlier than the ninth century.

When the Yenisei inscriptions were first discovered the alphabet in which they were written had not yet been decyphered. Some of them were copied by hand, others were photographed, not very successfully, and from others squeezes were taken, all of which were more or less imperfect owing to the very rough nature of the stones on which they were inscribed. Some of these squeezes were "retouched," that is an attempt was made to ink in the letters and this was often done incorrectly. It should be added that at least one of the most recent photographs, that of *Malov* 1952, No. 49, is not a photograph of the stone itself but a photograph of the stone after the letters had been chalked in, and careful scrutiny of the photograph shows that some of these chalk marks do not exactly follow the original letters. Thus it is hardly too much to say that, with very few exceptions, none of the hand copies were made by people who could read the texts which they were copying, and none of the editors had actually seen the original inscriptions which they were editing. The results have in some cases been disastrous. Study of the squeezes and the texts seem to show that in at least four cases, *Malov* 1952, Nos. 10, 24, 32, which has already been mentioned, and 39, and perhaps also No. 28, the stone was re-used and now bears two inscriptions which have been edited as if they were one. In a good many other cases the lines have been printed in the wrong order. It can be taken for granted that if an inscription contains a phrase like **er atım** "my name in manhood" or **atım** "my name" at or near the beginning of a line that line is the first line of the inscription. Thus for example in *Malov* 1952, No. 1, containing two lines, the inscription begins with the line printed second. In No. 25 of the three lines on the front of the monument the middle one comes first. Perhaps the most puzzling inscription of all is No. 41. It has an excellent "archaeological passport" (to borrow the expressive Russian phrase). It is in the Minussinsk Museum, and Malov actually saw and copied it, and yet it contains a number of letters unknown elsewhere and makes no sense

at all. The third line of five on the front of the stone begins **er atım** and is presumably the beginning of the text, but even the rest of this line is incoherent. It can hardly be a modern forgery, and there seems to be only two possible explanations. Either the mason who carved the inscription was so incompetent that he could not copy correctly the manuscript which was given to him or the man who drafted it was a fraud who visited tribal chiefs and drafted what purported to be funerary inscriptions which they themselves, being unable to read, did not realize to be mere collections of misshapen and meaningless letters.

(c) Five monuments, presumably funerary, found in the basin of the River Talas in Kirgizia. They are extremely rough and unsightly and largely unintelligible. In script and appearance they resemble the Yenisei inscriptions, and like them have generally been ascribed to an early date, fifth or sixth century. But as Kyzlasov (*op. cit.* footnote 18) remarks, they still have to be precisely dated. They carry no "archaeological passports," and unless some expert can examine them, like Kyzlasov, in their archaeological context, if this is still possible (one at least has been removed to a museum), they must remain undated but probably late.

(d) A number of texts written on paper, including the well-known *Irk Bitig* ("Book of Divination"), the longest text in Runic script, the letter signed **Bağa:tu:r Çigşi:** mentioned in Chapter I, and other miscellanea. Most of them have a Manichaean context, and are important because they contain a few Iranian words and phrases, which give known sounds for some of the letters. None of them has a precise "archaeological passport"; it is generally accepted that the likeliest date for them is about the ninth century, but their exact dating is not of great significance.

The Runic alphabet is in a class by itself. It is the only one of these alphabets which was invented, and used, only for writing Turkish. When the Runic inscriptions were first discovered, no one could read them, and V. Thomsen's decypherment of them was an intellectual feat of great

distinction. Unfortunately his decypherment was in essence an operation of pure deduction derived from an internal analysis of the texts. He did not consider how the letters of other contemporary alphabets which might have served (and at least one of which actually did serve) as a model for the inventor were used, and in particular whether any of the letters of such alphabets were polyphonic, and whether the letters used to represent vowels were used to indicate the quality or the length of the vowels represented or both. His decypherment was therefore based on two unstated principles:—

(1) except for the vowel letters, no letter was regarded as polyphonic;

(2) it was assumed that the vowel letters indicated only the quality of the vowel and not its length.

These two principles have been accepted without question by later writers on the subject, but, as will appear later, there is good reason for supposing that neither of them is valid.

The alphabet, for all its defects, is one of the most ingeniously devised alphabets of the first millennium, and it is obvious that the man who invented it had high intellectual and educational qualifications. It is not easy to guess in what milieu he made his invention, but I suggested some years ago, and still think, that the most plausible theory is that when the great Türkü ruler of the second half of the sixth century, Eştemi Kağan (the exact pronunciation of his name is uncertain, the Byzantines called him *Stembis Xagan*, and the Chinese *Shih-tieh-mi*) came into contact with civilized states in the West, the Sassanian and Byzantine Empires, he realized how greatly his administration of his broad domains would be facilitated, if he could, like the rulers of those Empires, communicate in writing with his subordinates, and gave orders for an official Turkish alphabet to be devised. This theory is not a new one, it was considered over sixty years ago by the pioneer of Runic epigraphy, Otto Donner, in his article *Sur l'origine*

de l'alphabet Turc, Journal de la Société Finno-Ougrienne XIV, 1, Helsingfors, 1896, but was discarded by him, chiefly because he believed, as it now turns out falsely, that the Yenisei inscriptions were the oldest examples of this script, in favour of the implausible theory that it was invented in the fifth or sixth century in some remote tribal area.

It is obvious that the inventor took as his principal model some form of the late Semitic (Aramaic) alphabet which had been adapted for use in writing some Iranian language. This suggests that he was inspired chiefly by the Chancellery practices of the Sassanian Empire; but it is not yet possible to pin down precisely the alphabet which was his immediate model. When Donner wrote his article the study of Iranian epigraphy was still in its infancy, and many of the most important monuments had not yet been rediscovered. There is still, so far as I am aware, no convenient comparative table showing the various forms which the Aramaic alphabet took in various places and at various dates, and if one looks at the tables which have been published, starting with that in Donner's article, the position is most confusing, since one Runic letter seems to find its best analogy in one alphabet and another letter in quite a different one.

But even if it be granted that one such alphabet was the principal model, it is obvious that this was no more than a beginning. Apart from inventing a number of letters for which no Aramaic model can be found, the inventor made three great innovations which distinguish his alphabet from the Aramaic/Iranian model:—

(1) he invented special letters used only to represent vowels—all the Aramaic letters in principle represented only consonants, but three of them were used to represent vowels also;

(2) he invented a number of digraphs, that is letters representing either a consonant preceded or followed by a specific vowel or vowels or two consonants without an

intervening vowel—he may have got this idea from the ligatures in Greek cursive script;

(3) he invented a number of letters which could be used only in words containing back vowels and a number of other letters which could be used only in words containing front vowels.

Before describing the Runic alphabet it is necessary to say something about the Aramaic/Iranian alphabet. It contained twenty-two letters, which can most conveniently be called by their Semitic names in the spelling of the Authorized Version of the English Bible, since these spellings are sufficiently eccentric and distinctive to make it clear without further definition that when these names are used it is letters of this alphabet that are being referred to. The names are:—*aleph*, *beth*, *gimel*, *daleth*, *he*, *vau*, *zain*, *cheth*, *teth*, *jod*, *caph*, *lamed*, *mem*, *nun*, *samech*, *ain*, *pe*, *tzaddi*, *koph*, *resh*, *schin*, *tau*. The alphabet was not wholly satisfactory even for writing Aramaic. It made no provision for writing vowels; originally these all had to be supplied to the reader, but by the time the alphabet was adapted for writing Iranian languages a practice had grown up whereby three of them, as well as being used to represent consonants, were also in certain circumstances used to represent long vowels. Furthermore there were of course far more than twenty-two sounds in the Aramaic language, so that a good many of the letters were used to represent more than one sound.

The following is a tentative table of the sounds (represented by letters with the phonetic values attributed to them in Chapter III) which were represented by the letters of the Aramaic alphabet at the time when it was adapted for writing Iranian languages:—

aleph	glottal stop; **a:**	*vau*	**w**; **o:**; **u:**
beth	**b**; **v**	*zain*	**z**
gimel	**g**, and perhaps **ǧ**	*cheth*	**h** or **x** or both?
daleth	**d**; **ḏ**	*teth*	**ṭ**
he	**h**	*jod*	**y**; **é:**; **i:**

caph	post-palatal **k**	*pe*	**p, f**
lamed	**l**	*tzaddi*	**ṣ**
mem	**m**	*koph*	velar **k**
nun	**n**	*resh*	**r**
samech	**s**	*schin*	**ṣ**
ain	' (as Arabic '*ayn*)	*tau*	**t; ṯ.**

The way in which this alphabet was adapted for writing Iranian languages has been so well explained by W. B. Henning in his article *Mitteliranisch* in *Handbuch der Orientalistik*, I, iv, 1 *Iranistik, Linguistik*, Leiden, 1958, that it is unnecessary to cover the whole ground here, but briefly the alphabet was an even more unsatisfactory instrument for writing an Iranian language than it had been for writing Aramaic. There were some Iranian sounds which did not exist in Aramaic; conversely some of the letters represented sounds which did not exist in Iranian. The actual process by which the Aramaic script was adapted for writing Iranian languages was a very unusual one, though not unique; much the same thing happened when the Sumerian script was adapted for writing Babylonian and the Chinese script for writing Japanese. The components used in the first instance for writing Iranian were not only individual letters but also whole words. For example, if what had to be written was "King Ardashēr," the proper name was spelt out in letters, but "king" was written with the Aramaic word MLK', although if the phrase was read aloud what was actually pronounced was the appropriate word for "king" in the particular language concerned, *ṣah* or some cognate form. In the initial stage, a whole short sentence, say a coin legend, apart from any proper names contained in it, could be written in these Aramaic words, which in this context are called "cryptograms" or "ideograms," so that it is sometimes quite uncertain whether such a coin legend is to be read as Aramaic or Iranian. Before long a number of Iranian words, as well as proper names, were spelt out in letters and so the proportion of cryptograms in the

sentence gradually fell, but the whole of the alphabet remained in use since every letter occurred in one cryptogram or another. On the other hand, with one exception, the letters which represented sounds unknown in Iranian, *teth*, *ain*, *tzaddi* and *koph*, were never used in spelling Iranian words. The exception was *tzaddi* which was taken into use in some Iranian languages (but not Parthian) to represent the unvoiced denti-palatal affricate **ç**. This reduced the effective alphabet to nineteen letters. There can be no reasonable doubt that when the inventor of the Runic alphabet took over such letters of the Aramaic/Iranian alphabet as he thought that he could conveniently use to represent Turkish sounds, he took them over to represent the full range of sounds, if there were more than one, which they then represented. Thus he took over *beth* to represent both **b** and **v**, *daleth* to represent both **d** and **d̲**, and so on. One proof of this, if proof be needed, is the fact that the letter used to represent **v** in the Iranian words *vre:şt* " angel " and *avre:n* " praise " in a Manichaean text in Orkun *op. cit.* II, 177, is the Runic equivalent of *beth* and that the names Rome (for Byzantium, with a curious prosthetic labial fricative) and Afar (Avar) in *KTE*.4 are spelt **Forom** and **Afar** with the **f** represented by the Runic equivalent of *pe*. The failure to recognize this fact has led, as I pointed out in *Turkish Ghost Words*, to some erroneous ideas about the phonetic structure of Türkü; for example it is often said that in that language " water " (*samech vau beth*) was pronounced *sub* when in fact it was pronounced *suv*.

Apart from the digraphs, the Runic alphabet contained thirty-two letters, five of them representing vowels:—**a/e**, **é** (only in the Tuvan inscriptions), **ı/i** (and **é** except in Tuva), **o/u** and **ö/ü**; and twenty-seven representing consonants. Of these twenty-seven, six represented **ç**, **m**, **ŋ**, **ñ**, **p/f** and **z** irrespective of the vowels of the words in which they occurred. One represented both **s** and **ş** in words containing front vowels. Two represented **s** and **ş** respectively in words containing back vowels, but these two letters

are very much alike and are liable to be confused; in fact they seem sometimes to have been confused by the masons who carved the inscriptions. The remaining eighteen letters can be arranged in nine pairs, the first of each used only in words containing back vowels and the second only in words containing front vowels. These pairs represented **b/v, d/d̲, t, ğ** and **g, k, l, n, r** and **y**. The number of digraphs is slightly uncertain; the ones in common use were those representing **ık/kı, ok/uk/ko/ku, ök/ük/kö/kü, lt** (or **ld** or both?), **nç,** and **nd** (or **nt** or both?); one representing **ıç/iç/çı/çi** was rather rare; a digraph representing perhaps the whole range **op/up/öp/üp/po/pu/pö/pü** and another perhaps with a similar range of vowel sounds but with **t** instead of **p** occurred occasionally in the manuscripts; there is a letter which seems to be a digraph and may be identical with the second of these in the Ongin inscription, see *The Ongin Inscription*, p. 185.

The forms of these letters and digraphs were fairly stable in the Orkhon inscriptions, and with minor variations in the manuscripts, but the Yenisei inscriptions contain some divergent forms, some of them very divergent, of these letters, but not apparently any additional ones except that representing **é**.

It should be added that there is one "ghost" Runic letter. A letter looking like a double axe has been read in *Toñ.* 26 and in *Malov*, 1952, No. 2, line 5 and 49, line 2. It has been explained as a digraph for **baş** by some scholars, but this is certainly an error. It is not a digraph but a letter and a digraph run into one either by the mason or by the copyist, in the first case **l ık** and in the other two either that or **p ık**.

The most readily available table of this alphabet is in von Gabain, *ATG*, p. 12. The first, and so far as I know the only, attempt to identify the Aramaic/Iranian originals of certain Runic letters was made in Donner *op. cit.*. Reviewing the matter in the light of later knowledge, we can say with some confidence that Runic **a/e** reproduces *aleph*, back **b/v** *beth*, **o/u** *vau*, front **s/ş** *zain*, velar **k** *cheth*,

post-palatal **k** *caph*, back **l** *lamed*, **m** *mem*, back **n** *nun*, back **s** *samech*, **p/f** *pe*, **ç** *tzaddi*, back **r** *resh*, back **ş** *schin* and front **t** *tau*. It should be added that it is possible that both back and front **r** represent different forms of *resh*. Back **r** is very like the ordinary form, but in Donner *op. cit.*, p. 24, there is a drawing of a supposedly indecypherable coin legend from Sogdiana, which when turned upside down looks like the Iranian name *ky'mrt* " Kayāmarṯ " with a *resh* very like a front **r**. This leaves *gimel*, *daleth*, *he*, *teth*, *jod*, *ain* and *koph* unaccounted for. No form of *gimel* is very like any Runic letter, but it might be reproduced by **ğ**; Donner was certainly wrong in identifying it with **ŋ** and probably wrong in identifying **ğ** with *he*. In fact the Aramaic/Iranian alphabet used by the inventor may have been one of those which did not include *he*. It is possible that back **d/ḏ** reproduces *daleth* but none of the forms are very close to the Runic letter, and there is one Iranian form of *teth* which is a little nearer. On the other hand, other forms of *teth* look more like back **t**. *Jod* was certainly used, but it is an open question whether it is reproduced by **ı/i** or front **y**; no form is like back **y**. No Runic letter bears any resemblance to *ain* or *koph* and it is unlikely that the alphabet which was used as a model contained either of these letters. It is therefore probable that the inventor used nineteen of the twenty-two Aramaic/Iranian letters as models for his alphabet. Conversely, of the thirty-two letters of the Runic alphabet, fifteen can certainly, and up to another five possibly, be identified with their Aramaic/Iranian models; this leaves at least twelve and perhaps more to be explained in some other way. One possible explanation is that they reproduced letters of the cursive Greek alphabet used by the Byzantines which was later to be the model for the Slavonic Glagolitic alphabet; there is no resemblance whatever between the unexplained Runic letters and the uncial Greek alphabet which was later to be the model of the Slavonic Cyrillic alphabet. Another, perhaps more plausible, explanation is that some of them reproduced letters not so much of the Byzantine cursive alphabet as of

the very much abraded Greek alphabet inherited by Eştemi's contemporaries and enemies the Hephthalites from the Kushans, and ultimately from the Graeco-Bactrians. A convenient table of the Kushan/Hephthalite alphabet will be found in R. Ghirshman, *Les Chionites-Hephthalites*, Cairo, 1948, p. 63. There is a fairly close resemblance between front **b** and *beta* in this alphabet, and between **z** and *zeta*; and *upsilon* is rather like a reversed **ö/ü**. If front **y** reproduces *jod*, then **ı/i** might be taken to reproduce *iota*. There is no resemblance between the remaining Runic letters and homophonic Kushan/Hephthalite ones, but front **d/d̲** has the same shape as some forms of *kappa* and *khi*, which are almost identical, and front **r** is very much like one shape of the Kushan/Hephthalite *tau*. It is conceivable that the inventor when he had run out of homophonic models for his alphabet in the alphabets which he used as models picked out one or two Kushan/Hephthanlite letters the shapes of which pleased him and used them to represent quite different sounds. The letter for **ñ** looks like two back **n**'s intertwined, but this may be a mere coincidence. After all these possibilities have been taken into account, quite a number of letters, and with one possible exception all the digraphs, remain unexplained; there is no resemblance between any of the digraphs and the letters representing their constituent sounds, which casts some doubt on the theory that they were modelled on the Greek ligatures. It has, however, been suggested rather plausibly that **ok/uk/ko/ku** which looks like an arrow-head standing on its point is a rebus for **ok** "arrow."

There must have been some recognized alphabetic order for the Runic letters, but it has not survived. Von le Coq found a scrap of paper at Turfan containing part of a list of Runic letters with the equivalent sounds in Manichaean Syriac script, the most accessible reproduction of which is in Orkun *op. cit.* II, 24. Only nineteen entries have survived; the letters are in no recognizable order, ordinary letters and digraphs being jumbled up indiscriminately, and the

transcriptions are so erratic as to suggest that the writer was not very familiar with the Runic alphabet.

There were clearly strict orthographic rules for the use of the alphabet, which were more or less faithfully observed in the Orkhon inscriptions, less faithfully observed in the manuscripts, and largely disregarded in the Yenisei and Talas inscriptions. Four of these rules can be identified:—

(1) Letters specially designated for use exclusively in words containing back or front vowels respectively should be used only in such words. There are only one or two departures from this rule in the Orkhon inscriptions (chiefly in *Toñ.* and the Ongin inscription) and the manuscripts, but there seem to be a good many more in the Yenisei inscriptions. Not all the apparent transgressions of this rule, however, are actual. Grønbech has suggested that some Türkü affixes, for example the verbal affix *-miş*, were invariable, irrespective of the vowels in the word to which they were attached. Thus a spelling like **sokusmi:ş** in *IB*, para. 2, is more likely to be an accurate representation of the sound of the word than an orthographic error.

(2) Double consonants should be written with a single letter. As shown in Chapter VI, double consonants within a basic word are extremely rare, but they occur more often when a suffix starting with a consonant is attached to a basic word ending in a homorganic consonant. In both cases only one letter was used; thus **ékki:** "two" was written *é k i* or *k i* and **ıtti:** (for **ıdtı:**) "he sent" *ı t ı*.

(3) Short vowels, other than those enclosed in digraphs, should not be written except when they are the first vowel of a word, and then only if they are not **a/e**. This rule was so strict that even an initial short **a** was left unwritten, so that for example **at** "horse" was written *t*. Conversely vowels other than **a/e** should be written, even when short, if they are the first vowel of a word. Attempts have sometimes been made to transcribe Runic texts on the principle that short vowels other than **a/e** can be assumed in the first syllable even when not written. There are one or two

faulty scriptions of this kind, for example in the Ongin inscription; but generally speaking such assumptions are false. The practice in regard to long vowels was not consistent; this seems to be an inheritance from Aramaic/Iranian orthography, where also the practice was not consistent. In principle long vowels should always be written; in practice, except sometimes in the Yenisei inscriptions, final long vowels were always written. In the medial position long vowels at the end of open syllables were normally written, as in **kün ortu:sı:ŋaru:** *KTS*.2, but were sometimes omitted, as in **sü:ledim** (for **sü:le:dim**) *KTS*.3 *etc*. In closed syllables long vowels were sometimes written, as in **altu:n** *KTS*.5, but sometimes, indeed probably more often, omitted. This is particularly the case with **a:/e:**, which, though invariably written where representing a final long vowel, and usually, but not always, when representing an initial one (**a:t** " name " is normally spelt *a t*, but **a:v** " wild game " is invariably spelt *v*), or a long vowel in an open syllable, seems never to be written to represent a long vowel in a closed syllable (**ta:ğ** " mountain " is written **t ğ**). Looking at the matter from the opposite point of view, the vowel letter **a/e** invariably represents a long vowel, the other vowel letters in the second or a subsequent syllable invariably represent a long vowel, but in the first syllable they may represent either a long or short vowel, and evidence regarding the length of such a vowel must be sought elsewhere. This rule was applied with its full rigour only in the Orkhon inscriptions. The manuscripts were much freer with their use of vowel letters; there are cases where vowel letters, even **a/e**, are used to represent what were almost certainly short vowels but no obvious cases of vowel letters being absent when their presence might have been expected. In the Yenisei inscriptions the spelling is so chaotic that the term orthography hardly applies. There are cases where vowel letters represent short vowels and cases where even final long vowels are not represented by vowel letters. There is in all the texts some inconsistency in the use of vowel letters

in the first syllable in association with digraphs. Theoretically the digraph is sufficient by itself to indicate the nature of the vowel in the syllable and no further indication is necessary; O. N. Tuna, in his article *Köktürk ve Uygurcada Uzun Vokaller*, Türk Dili Araştırmaları Yıllığı, 1960, has argued that spellings like *ko o ñ* " sheep " and *ku u t* " good fortune " necessarily imply that these words were pronounced **ko:ñ** and **ku:t,** but Kāşğarī's spellings are **ko:n/ko:y** and **kut,** and these are the spellings which we should expect, since long vowels do not occur in monosyllables ending in an unvoiced plosive, but are common in monosyllables ending in a nasal. Moreover in *KT* the proper name (or title?) **Kül** is usually spelt *k ü l*, once or perhaps twice *kü l* and once *kü ü l*. Thus it seems clear that a vowel letter other than **a/e** in the first syllable is no indication of length even when it is associated with a digraph.

This is perhaps the most convenient place at which to point out that the loss of the special letter for **é** caused great inconvenience to the writers of Runic script. They never made up their minds whether to regard it as a sound close to **i** and so to be represented by the same vowel letter as **i,** or as a sound close to **e** and so to be left unrepresented; very often a word like **él** " realm " containing this sound is spelt with the vowel letter in one place and without it in another place in the same inscription, or, in the case of passages which appear both in *KT* and *BK*, with it in one inscription and without it in the other.

(4) Finally there were rather elaborate and not entirely clear rules relating to punctuation. Runic texts were not punctuated in the modern sense of that term, but they were broken up into very short sections by a mark resembling a colon (:). In the Orkhon inscriptions these sections are normally single words, occasionally two and very occasionally three. Thus for example in the first line of *KT* the opening words **Teŋri: teg** " God-like " are followed by a colon; the next seventeen words are all separated by

colons; then the words **Şad̲apı:t begler** are placed between colons; then come a number of single words all separated by colons; these are followed by three phrases **tokku:z oğuz; begleri: bod̲unı:** and **bu: savımın** each between a pair of colons; and these are followed by a number of single words separated by colons. So far as it is possible to judge, having regard to the extreme roughness of the stones, the practice is much the same in the Yenisei inscriptions, but it is difficult to be certain. In the manuscripts the practice of separating each word by a colon from the next seems to be universal. In the Talas inscriptions there are no colons at all, but it rather looks as if in these and perhaps some Yenisei inscriptions the vowel letter **a/e** was used instead of a colon. It is difficult to determine what principle the draftsmen followed in their use of colons, but it is possible that they put two or three words between a pair of colons for much the same reasons that we join two or three words with hyphens. **Teŋri: teg:** and **tokku:z oğuz:** are exact analogues of our "God-like" and "Tokkuz-Oğuz," but we should not normally join "*begs* and people" or "these words of mine" with hyphens.

It will be remembered that medial short vowels are not represented by a vowel letter; the same practice seems to have been followed in the case of final short vowels at the end of the first word of a pair of words enclosed by colons, but there seem to have been very few words in early Turkish which ended in a short vowel. By the eleventh century all final vowels were long; there may in a much earlier stage of the language have been quite a number of words ending in short vowels, particularly perhaps after two consonants, but even by the eighth century they seem to have been very rare. The only certain case is the tribal name **Türkü.** By the eleventh century, and perhaps even earlier, this had admittedly become **Türk,** but there does not seem to be any doubt that in the Runic texts it was **Türkü** normally spelt **t ü r kü.** For some unknown reason the editors have been unwilling to transcribe this in the

obvious way and have habitually transcribed it *Türük*, sometimes adding that this should be pronounced *Türk*. There does not however seem to be any justification for this practice, once it is realized that even if the name is spelt *türk* as the first of a pair of words enclosed by colons the pronunciation *türkü* is as possible as *türk*. The facts are these:—

(a) The word occurs seventeen times in *Toñ*. the oldest long inscription which has survived and probably the oldest inscription of all. In lines 1, 2, 9, 17, 18 it occurs, followed by **bo<u>d</u>un** in a phrase enclosed by colons and spelt : *türkbo<u>d</u>n* :; in line 3 it occurs in the phrase : *türk* : *şirbo<u>d</u>n* : and in line 11 in the same phrase but without the medial colon. In line 20 it occurs in the phrase : *õŋretürkkğnğru* :. Thus up to this point in the inscription the word is spelt *türk*, which is only once followed by a colon, if indeed this medial colon in line 3 is not a scribal error. The word does not occur again until line 46, where the spelling is : *türkübo<u>d</u>n* :. It occurs eight more times—in lines 50 (twice), 54, 58, 60, 61 and 62 (twice)—always in the same spelling and always as the first word in a group. In three of these occurrences—lines 60, 61, 62—it occurs in the phrase : *türkü şirbo<u>d</u>n* :. The only possible explanation seems to be that Toñukuk, who drafted his own inscription, started by spelling the word *türk* because he thought that readers would naturally supply the final short vowel, but later realized that the use of the digraph would be more explicit. As the variations already mentioned in the spelling of **Kül** show, the draftsmen never seem to have made as much use of this digraph as they might have.

Incidentally the passages quoted above are the only ones in which the word which has usually been transcribed *sir* or *şir* occur. I suggested to Dr. Aalto (see p. 50 of his edition of *Toñ*.) that this word might be taken as a loan word from Sanskrit *śrī* " auspicious " and transcribed **şiri**. If so this is another example of an unwritten short vowel.

(b) The word occurs nearly fifty times in *KT* and *BK*, counting an occurrence in a passage common to both inscriptions as one only, almost always as a single word between colons, and invariably spelt *t ü r kü*.

(c) It next appears in the Ongin inscription, which, as I pointed out in my recent edition, is a pastiche of *Toñ.* and *KT*. It occurs three times in *Toñ.*'s earlier spelling in the phrase : *t ü r k b o d n* : surrounded by colons.

(d) It appears twice in *KÇ*, which also seems to be largely a pastiche of *Toñ.* and *KT*, on both occasions in the phrase : *t ü r k b o d n* : surrounded by colons.

(e) In the oldest Uyğur inscription, *Şu*, the funerary memorial of the first Uyğur *kağan*, it occurs four times, always spelt *t ü r kü* and usually as the first word of a phrase enclosed by colons. Incidentally this is the only Turkish text in which the **Türkü bodun** are mentioned as enemies who were " destroyed."

(f) The only other supposed occurrence of the word is in *Malov*, 1952, No. 32, but for the reasons given above no reliance can be placed on existing editions of that text.

If these scriptions stood by themselves, it could perhaps be argued that, since the word was undoubtedly **Türk** in the eleventh century and later, *t ü r kü* must be taken as an eccentric and inexplicable scription for the same word. But they do not; foreign scriptions of the name all support the theory that it had a final vowel:—

(a) In the Saka (Khotanese) documents, which are now dated to the ninth and tenth centuries, that is to a period after the downfall of the Türkü Empire, but one in which the name was still current as a tribal appellation, it is spelt *ttūrki*, *ttrūka*, *ttrūki* (H. W. Bailey in J.R.A.S., 1939, p. 88) and *ttrūkä* (ditto in Asia Major, New Series I, 1, p. 29*).

(b) The normal spelling in early Tibetan documents relating to events in the seventh and eighth centuries was

* In the Brāhmī/Saka alphabet t was normally represented by *tt*; *t* seems to have been pronounced ṭ.

Dru. gu (with some variations); see F. W. Thomas, *Tibetan texts and Documents*, Vol. 2, London, 1951, pp. 267 ff.

(c) The evidence of the Arabic historians and geographers is not conclusive since a final short vowel would have been represented by a *ḍamma*, and *Turku* might have been declined like any ordinary Arabic common noun ending in *-u*.

(d) Finally we come to the Chinese scription of this name, *t'u-küeh*, in Karlgren's "Ancient Chinese" *tuɔt-kiwɔt*. The interpretation of this scription has been bedevilled by the unfortunate *obiter dictum* of the late Prof. P. Pelliot in his article *L'origine de T'ou-kiue, nom Chinois des Turcs*, T'oung Pao, 1915, p. 687, "The Chinese name *T'ou-kiue* must represent a Mongolian (Juan-juan) plural, Türküt, of **türk**, literally 'strong'." It will be noted that the implication could be taken to be that the scription represents not what the Türkü called themselves, but what the Juan-juan called them, but, considering how intimate the relations between the Chinese and the Türkü were, it is not very probable that the Chinese should have gone to the Juan-juan for a name for them.

On one point at least Pelliot is certainly wrong. In an article, "*The concept of 'strength' in Turkish*," awaiting publication by the Türk Dil Kurumu, I have shown that **türk** does not mean "strength" but "ripeness" (of fruit), "maturity" (of a man) and the like, and that there is no good reason for supposing that there is any connection between the word **türk** and **Türkü**. Moreover, I suggested in Chapter I that the Juan-juan were Hsien-pei and not Mongols, and so probably spoke a form of **l/r** Turkish. This does not, it is true, dispose of the possibility that **-t** might be a plural suffix, since there are traces of an obsolete plural suffix **-t** in Turkish, which probably has no connection with the Mongolian plural suffix *-d* (*sic*, not *-t*). The real difficulty about Pelliot's theory is that in all the Chinese scriptions of Turkish names and words in the T'ang period there is not a single trace of a scription in

which "Ancient Chinese" *-t* represents a Turkish **-t**; in fact when a Turkish **-t** at the end of a syllable had to be represented an extra syllable with initial *t-* was inserted; for example **kutluğ** was transcribed *ku-tu-lu*, not *ku-lu*, even though the "Ancient Chinese" pronunciation of *ku* was *kuət*. There are convenient collections of Chinese scriptions of Turkish names and words with the original characters in Chavannes *op. cit.*, Liu *op. cit.* and Hamilton *op. cit.* Careful examination of these scriptions shows that the only Turkish sounds which are represented by "Ancient Chinese" *-t* are **d̲**, **l**, **r** and *zero*. A good many illustrations of each could be given, but the following specimens will suffice:—

d̲	*sha/shê/ch'a* (*ṣat/ṣiat/tṣ'at*)*	**ṣad̲**
l (and *zero*)	*pa-ssŭ-mi* (*b'wat-siet-miet*)	**Basmıl**
r *mo-ho-tu*	(*muo-γâ-tuət*)	**Bağa:tu:r**
zero (and **r**)	*kü-lü-ch'o* (*k'iuət-liuət t'śiwät*)	**Küli Çor**

Prof. Pulleyblank, although he fully agrees that *-t* represents *zero* within a word, as in **Basmıl**, is unhappy at the suggestion that it should be entirely functionless at the end of a word, since the Chinese could have found scriptions with vocalic endings to represent Turkish final syllables with vocalic endings. But the fact remains that *kiuət-liuət* cannot conceivably represent anything except **küli** (or **külü**); and the odds are heavily against *tuət-kiwvt* representing anything except **Türkü.**

In view of the convergence of the Runic and foreign scriptions it does not seem to me possible to resist the conclusion that the original form of the name was **Türkü** with a short final vowel.

Another possible example of a word ending with a short final vowel is the proper name (or title) **Küli.** *KÇ*, which is much damaged, seems to be the memorial to a man called **Küli Çor.** The text has been edited from an indifferent squeeze, and it is very likely that even the edition in *Malov*, 1959, is incorrect in points of detail, more particu-

* The main entries are in Wade transcriptions, those in brackets are "Ancient Chinese."

larly as regards the presence or absence of colons, but if it is to be relied on the name is normally spelt : *kü ü l i ç o r* :, but once : *k ü l* : *ç o r* :. This man is probably the same as the leader of the Tarduş begs at Bilge: Kağan's funeral, whose name is spelt : *k ü l ç o r* : in *BKS*.13. The same name, and perhaps the same man, appears in the first Xoyto-tamir inscription (*Malov*, 1959, p. 47) as *k ü l ç o r*. Practically no colons can be seen in these inscriptions, which were painted and not engraved. The *T'ung Tien* (see Liu *op. cit.*, p. 498) in a list of Turkish titles includes *k'ü-lü ch'o* (*k'ɨuət-lɨuət t'ɨwät*), which, as stated above, must represent **küli** (or **külü**) **çor.** There does not, therefore, seem to be much doubt that this word too had a short final vowel, since a long final vowel could hardly have been omitted. This suggests the intriguing possibility that the name Kül Tégin, which is invariably spelt : *k ü l* (or *kü l*, or *kü ü l*) *t é g n* : between two colons, should properly be transcribed Küli Tégin, but against this must be set the fact that the Chinese scription of this name was *k'üeh* (*k'ɨwăt*) *t'ê-k'in*, which points clearly to **Kül Tégin.**

To sum up, the Runic alphabet possessed letters (apart from digraphs which are merely combinations of letters) which had the following phonetic values:—

VOWELS. **a:/e:, é/é:, é/é:/ı/ı:/i/i:, o/o:/u/u:, ö/ö:/ü/ü:.**

It is unfortunate that the special letter for **é/é:** occurs only in the Tuvan inscriptions, the vocabulary of which is very limited. In view of its importance as indicating in what words this sound existed, it may be useful to give a list of the words in which it occurs. The references are to the inscriptions in *Malov*, 1952.

ék(k)i:	" two," 49, 3.
él	" realm," nearly twenty occurrences.
él(l)e:-	" to incorporate in the realm," 45, 4.
élçi:	" envoy, ambassador," 1, 2; 14, 1.
él(l)ig	" fifty," 15, 3 (or perhaps " having a realm "); 49, 4.
éş	" comrade," 2, 1 (doubtful).

bél	"waist," 3, 2; 10, 5.
béş	"five," 25, 4; 45, 2.
kéş	"quiver," 3, 2; 10, 5.
yétmiş	"seventy," 3, 4; 45, 5.
yégi:rmi:	"twenty," 45, 2.
yér	"place, country," 45, 6.

The weak points in the alphabet were:—(1) many short vowels and some long ones are unwritten and have to be inferred; (2) there is no means of telling whether a vowel in the first syllable, other than **a/e**, is short or long; (3) except in the Tuvan inscriptions there is no means of telling whether a vowel is **é** or **i**; (4) there is no means of distinguishing **o** from **u** and **ö** from **ü**. The strong point is that owing to the double representation of some consonants it is nearly always possible to tell whether a word contains back or front vowels, and there is a fairly liberal supply of vowel letters, particularly in the first syllable.

CONSONANTS. **b/v***, **ç**, **d/d̲***, **t***, **ğ**, *velar* **k**, **g**, *post-palatal* **k**, **l***, **m**, **n***, **ŋ**, **ñ**, **p/f**, **r***, *back* **s**, *back* **ş**, *front* **s/ş**, **y***, **z**.

Those marked with an asterisk (*) were in two forms used respectively in words containing back and front vowels.

It is possible, but improbable, that the letter which represented **ç** also represented **c**. **Ğ** was probably a fricative, but may sometimes have been a plosive; **g** was probably a plosive but may sometimes have been a fricative. The letter which represented the velar plosive **k** may also have represented the homorganic fricative **x**, but it is uncertain whether this sound existed in Türkü.

The only weak point of the alphabet, inherited from its Aramaic/Iranian model, is the use of the same letter to represent plosives and homorganic fricatives. Its most important feature is the existence of a special letter for the palatal nasal **ñ**. The sound was a rare one in the eighth century and probably always had been; its subsequent evolution was into **n**, **y** or **yVn** according to the language concerned. Its existence in a number of words can be

proved indirectly by comparing forms of the same word containing different derived sounds in different languages and by noting words with an initial **m-** representing an original **b-** which had been nasalized by regressive assimilation, see *The initial labial sounds in the Turkish languages*. But the Türkü texts provide the only direct evidence of its existence, and it may be useful to give a list of the words in which it is found, with references to the texts in which they occur.

añığ	" bad, excessive(ly) " *Toñ.*, *KT*, *BK*; *IB*, *MSS*.
çığañ	" destitute " *KT*, *BK*; *IB*.
turñya: (*sic*?)	" a crane " *IB para.* 61.
kañu:	" which?" *Kara Balgasun inscription* (Orkun *op. cit.* II, 38; C.3).
koñ	" sheep " *KT*, *BK*; *Şu*; *IB*, *MS* (Orkun *op. cit.* II, 96).
koñçı:	" shepherd " *MS* (Orkun *op. cit.* II, 67).
koñlığ	" owning sheep " *Şu N.*6 (reading uncertain).
yañ-	" to disperse " *Ton.*, *KT*, *BK*.

Proper names:—

Toñukuk	*Toñ.*, *KÇ*.
Kıtañ	*Toñ.*, *KT*, *BK*, *KÇ*.

There does not seem to be any certain occurrence of ñ in the Yenisei inscriptions, although it has been read in one or two passages. It is a mere coincidence that all the words listed above contain back vowels. The sound also occurs in words containing front vowels, like * **béñi:** " brain."

(ii) THE TEXTS IN BRĀHMĪ SCRIPT

Parts of fifteen manuscripts containing Turkish texts written in Brāhmī script have survived, and have been admirably edited by Prof. von Gabain in *Türkische Turfantexte**, VIII, A.D.A.W., Berlin, 1954. All are more or less fragmentary. The first eight (A to H) are bi-linguals, that

* Hereafter cited as *TT*.

is Buddhist Sanskrit texts broken up into single words or short phrases, each followed by a Turkish translation; three (K, N and O) are Buddhist texts in Turkish; two (I and M) are medical texts, presumably translated directly or indirectly from Sanskrit; one (L) relates to astronomy and one (P) to the calendar. Reproductions of parts of C, F and G are included in *TT* VIII, and a reproduction of part of I, together with a table of the alphabet, including examples of vocalized letters and ligatures, in von Gabain *ATG*. A reproduction of part of D was published by H. Stönner in *Zentralasiatische Sanskrittexte in Brāhmīschrift aus Idikutšahri, Chinesisch Turkistan*, I, S.K.P.A.W., Berlin, 1904. There are some variations of script and much greater variations of orthography between the different manuscripts, but there is no reason to suppose that these variations are due to large differences in the age of the manuscripts. The actual date at which they (and the manuscripts from which they were copied, if these are not the prototypes) were written is quite uncertain. The Brāhmī alphabet was introduced from India into Central Asia at a "fairly early" date and was there used to write not only Sanskrit but also "Tokharian A and B," Saka and Turkish. The latest manuscripts in this script, for example the Staël-Holstein roll, were written as late as the tenth century, but the earliest must be a good deal older. Prof. von Gabain has told me that the quality of the paper on which the Turkish texts were written suggests that they are rather late, perhaps as late as the tenth century, and she is inclined to think that they are to be connected with an attempt to stimulate Turkish Buddhism in Turfan at about that time. But the first use of the Brāhmī alphabet to write Turkish must be dated to an earlier period. It will be shown later that the orthography of the Turkish texts in Tibetan script was based on the orthography of the Brāhmī texts, and the Tibetan manuscripts are certainly earlier than the tenth century. Prof. von Gabain has suggested that the differences between the orthography of the various manuscripts reflect differences of dialect, but

it seems to me that they are all written in Uyğur, with some traces of Uyğur-A in one or two texts, and that the differences are rather due to differences in the educational qualifications of the various scribes or perhaps their ethnic status; it is very probable that some, if not all, of them were not ethnically Turkish.

The Brāhmī alphabet, or rather the prototype from which it was derived, was by many centuries the earliest scientific alphabet to be invented, a beautifully precise instrument for recording the nuances of the pronunciation of Classical Sanskrit; but by the time that it reached Central Asia, some centuries after it had been invented, it was no longer such a perfect instrument. The reason for this was, quite simply, that by this time the pronunciation of Sanskrit, which was then no doubt a dead language, had altered a good deal although the spelling remained more or less "frozen." The numerous mistakes in spelling in the Sanskrit parts of the bilinguals, mainly but not exclusively in regard to the length of the vowels, show that the scribes were already in difficulties owing to the difference between current pronunciation and the traditional spelling. The spellings, and in particular the inconsistencies of spelling, of the Saka manuscripts too show that some of the consonantal sounds had changed; in particular it seems that there had been alterations of voicing and that some plosives had become fricatives. The subject is discussed in H. W. Bailey's "*Languages of the Saka*" in the same volume as W. B. Henning's *Mitteliranisch*, cited above.

Whatever the exact pronunciation of the Sanskrit letters when they first began to be used for writing Turkish, it was in any event clear that they were not by themselves sufficient to represent all the Turkish sounds. So far as the vowels were concerned, there was no difficulty about representing short and long **a**, **i** and **u**, although long **i:** was in fact seldom written, but the signs for closed **é** and **o** represented sounds which were assumed to be long and there was no means of representing short, as opposed to long, **é** and **o**. No way ever seems to have been devised

of distinguishing between back **ı** and front **i**. So far as the other front vowels, **e**, **ö** and **ü**, are concerned, the ingenious device was adopted of attaching a subscript *y* to the preceding consonant, so that, for example, **küse:yü:** in *TT* VIII A.1 was written *kyu syā yyu*. The system is only occasionally ambiguous, and then chiefly owing to the practice of dividing the Turkish phrases not into words but into convenient syllables. Thus, for example, von Gabain was technically justified in transcribing **ā śa ğyā rsi r** in *I*.8 **a:şağ e:rsir**, but what the scribe intended was **a:şağ ya:rsır** "he is revolted at (the sight of) food." It seems that some of the scribes had difficulty in distinguishing acoustically between **ö** and **ü**, so that words which were certainly pronounced, say, **ö:d̲** "time" and **öt** "advice" are sometimes written with *ü*-. There are also great inconsistencies in the vocalizations of the same word in different manuscripts and sometimes even in the same manuscript. This in particular applies to the length of vowels. For example Kāşğarī makes it clear that in Xākānī the plural suffix **-la:r/-le:r** always had a long vowel, and it seems reasonable to suppose that the pronunciation was the same in Uyğur. In the Brāhmī texts **-la:r/-le:r** is the standard spelling in A, C, F, G, H, I, K, L, M, O and P. The suffix does not occur in B, a very short text, but other spellings like **-ma:k** suggest that here too it would have been **-la:r/-le:r**. On the other hand, in N the spelling is invariably **-lar/-ler** and in D and E, in which Sanskrit long vowels are often written as short, the usual spelling is **-lar/-ler**, but **-la:r/le:r** occurs occasionally. It should be added that short or super-short vowels are sometimes omitted as, for example, in *ymyā* A.1 for **yeme:**.

To sum up, the system of notation of vowel sounds in the Brāhmī script was in theory an almost perfect instrument for representing Turkish vowel sounds; there was practically no polyphony and the only defects, minor ones, were that there was no means of distinguishing **ı** from **i** or between short and long **é**, **o** and **ö**. A long list of words containing closed **é** can be compiled from the index to

TT VIII, and one important point which the texts in Brāhmī, and in a minor way Tibetan, script disclose is that in Uyğur, unlike most Turkish languages, **o** and **ö** could occur in the second and subsequent syllables and not only in the first. In most words, for example **törö:** "unwritten law," the **o/ö** followed an **o/ö** in the first syllable, but in some, for example **alko:** "all," **artok** "more" and **egsö:-** "to be lacking" it followed **a/e.** In practice, however, the system, owing to the carelessness or lack of skill of the scribes, was not always satisfactory in indicating the vocalization of particular Turkish words.

So far as consonants are concerned the Brāhmī script was in theory an equally perfect instrument for representing Turkish sounds. Six, or perhaps seven, new letters (see the table in von Gabain *ATG*) were devised for representing sounds which could not conveniently be represented by the ordinary Brāhmī letters, and it is fairly easy to infer what sounds these letters, which included *ker* for velar **k** and *rr* for **ğ**, were intended to represent. But in practice the script is an even less satisfactory instrument for representing Turkish consonantal sounds than it is for representing vowel sounds. This is partly due, again, to the carelessness or lack of skill of the scribes, but even more to the fact that it is clear that by this time the elaborate phonetic notation of the Brāhmī alphabet had so far broken down that the scribes were quite uncertain what sounds certain letters represented, and were frankly at a loss to know which letter should be chosen to represent a particular Turkish sound. For example a careful analysis of the words beginning with a labial plosive shows that this sound was represented only by *p-* or *ph-* in A, B, C, E, G and N, only by *b-* or *bh-* in F, H and I, and by all four letters indiscriminately, usually with a fairly marked preference for one pair over the other, in D, K, L, M, O and P. The special letter *rr* which was invented to represent a voiced fricative seems to have been used indiscriminately to represent velar **ğ** in words with back vowels and post-palatal **g** in words with front vowels. This may be taken as indicating that the

voiced post-palatal **g** was, at any rate in certain contexts, a fricative sound. It is perhaps significant that the Sanskrit **g** was seldom, if ever, used. Some manuscripts show other idiosyncrasies in the representation of velar and post-palatal sounds. Some represent the velar **k** by *hk* instead of *kr*, some use *h* for **ğ** and some *k* for **g**. It is possible that a very close analysis of the spellings of various words in different manuscripts would bring out further points of interest, but it might turn out to throw more light on the evolution of the phonetic value of the Brāhmī letters than on the phonetic evolution of Turkish. Thus, while these texts give us important evidence regarding the vocalic structure of Uyğur, it would be rash to draw any inferences from them regarding the consonantal structure except perhaps in regard to the pronunciation of post-palatal **g**.

(iii) THE TEXTS IN TIBETAN SCRIPT

Only half a dozen fragments of Turkish texts in Tibetan script have survived, and these have not been published, but the late Prof. F. W. Thomas gave me photographs of what he believed to be the whole collection, and the following remarks are based on a study of them. The fragments comprise:—

(1) parts of lines 142 to 161 of the *Sekiz Yükmek Sutra*, published in *TT* VI, S.P.A.W., Berlin, 1934, and another small fragment probably of the same manuscript;

(2) a text starting with a series of *namaskara's*, that is invocations of various Bodhisattvas, similar to the extract from the Turkish version of the *Suvarṇaprabhāsasūtra* published in F. W. K. Müller, *Uigurica*, A.K.P.A.W., Berlin, 1908, pp. 17 ff. and with one or two identical sentences;

(3) parts of a short list of proper names, some Turkish and some Tibetan;

(4) a very small fragment of a Buddhist text.

The collection also included a small fragment of a Tibetan text containing a Turkish proper name (*'yel.grab.*

śud.myiś) and a quasi-Sanskrit *dhāraṇi* written out in parallel columns in Tibetan and Uyğur script.

These were all written in *dbu.can* script of varying degrees of formality, high in (1) and (4), medium in (3) and low in (2), and Prof. Thomas dated them all on palaeographic grounds as "rather early" by Tibetan standards, that is not earlier than the eighth century but not as late as the tenth. No very useful purpose would be served by publishing the texts *in extenso*, but the following list of the Tibetan scriptions of all the Turkish words which they contain, except one or two Sanskrit loan words, accompanied by a transcription in normal Uyğur, may be of interest. They are arranged in the order proposed in Chapter IV, declensional and other similar suffixes being disregarded. *SY*, *NK*, *PN* and *Fr.* indicate that the word has been taken from (1), (2), (3) or (4) respectively.

'a.tu.myi.ś	*PN*	**ata:mış** (?)
'yed.ku	*NK*	**ed̲gü:**
'u.du.nur	*SY*	**ud̲unur**
'yo.do.nu	*NK*	**ötönü:**
['*a*].*ḫda.sis*	*SY*	**ad̲asız**
'yo.tos	*NK*	**et'öz** (?)
'og	*SY*	**ok**
'uk.sar	*SY*	**uksar**
og.li	*SY*	**oğlı:**
'a.gri.lig.g yi	*NK*	**ağılığı:**
'a.gri[*r.la.*]*yur*	*SY*	**ağırlayur**
'yog.li	*NK*	**ögli:**
'yi.kyin.ti	*Fr.*	**ikinti:**
'yog.run.cu	*PN*	**ögrünçü:**
'yel (*grab.śut.myiś*)	*PN*	**él (kavşutmış)**
'yal.da	*NK*	**élde:**
'ol	*SY*	**ol**
'al.ti	*SY*	**altı:**
'u.la.ti	*SY*	**ula:tı:**
'u.lug	*NK*	**uluğ**
'al.gro	*SY*	**alko:**

'al.grin.sis	*SY*	**alkınçsız**
'ye.lyig.lig	*NK*	**éliglig**
'u.mug	*NK*	**umuğ**
'am.ri.lu[r]	*SY*	**amrılur**
'yoṅ	*SY*	**öŋ**
'yin.ca	*SY*	**ınça:**
'ye.rib	*SY*	**erip**
'yart.myiś	*NK*	**ertmiş**
'yer.tya.ṅu	*SY*	**ertiŋü:**
'a.rig	*PN*	**arığ**
'yarg.lig	*NK*	**erklig**
'aḫi	*Fr.*	**a:y**
'yi (tog.ril)	*PN*	**a:y (toğrıl)**
'a.[ya.]yur	*SY*	**aya:yur**
'u.zu.nin	*NK*	**uzunın**
ḫbyi.ti.kyig	*SY*	**bitigig**
pag.lig	*NK*	**bağlığ**
ḫbyi.lir	*SY*	**bilir**
ḫbyi.lig	*SY*	**bilig**
ḫbyil.kya	*SY*	**bilge:**
ḫbyi.lig.li[g]	*SY*	**biliglig**
byi.lig.lig	*NK*	**biliglig**
bol.myiś	*PN*	**bolmış**
bol.ci	*NK*	**bol[da]çı: (?)**
bur.han	*NK*	**burxan**
pars	*PN*	**bars**
ḫbyeś	*Fr.*	**béş**
paś.lag.gri.ni	*NK*	**başlağını:**
tu.byin	*NK*	**tübin**
tab.graḫ	*NK*	**tapğa:n (?)**
tyo.byon	*Fr.*	**töpön**
ta.pyi.nur	*SY*	**tapınur**
tub.rhag	*Fr.*	**toprak**
to.tog	*PN*	**tutuk**
tu.daḫ.si[s]	*SY*	**tudasız**
thug. (tog.ril)	*PN*	**tuğ (toğrıl)**
('yi). tog.ril	*PN*	**(ay) toğrıl**
(thug). tog.ril	*PN*	**(tuğ) toğrıl**

(*gut.lug*) *tog.ril*	*PN*	(kutluğ) toğrıl
(?) *tog.ril*	*PN*	(?) toğrıl
ti.kyin. (*kyol.miś*)	*PN*	tégin (külmiş)
te.ku.nur.myen	*NK*	teginürmen
din.lig	*SY*	tınlığ
tyo.rog	*SY*	törög
tyord	*Fr.*	tört
tur.gra.ru	*SY*	turkaru:
tyor.lu[*g.*]	*SY*	törlüg
tvoz	*SY*	töz
tvo.zun	*SY*	tözün (tözön)
tyo.zon	*NK*	tözün (tözön)
to.zon	*NK*	tözün (tözön)
('*yel*) *grab.śud.myiś*	*PN*	(él) kavşutmış
gu.ti.ńa	*NK*	kutıŋa:
gut.lug (*tog.rıl*)	*PN*	kutluğ (toğrıl)
grut.lug	*NK*	kutluğ
grut.ru.[*lur*]	*SY*	kutru[lur]
gral.myiś	*NK*	kalmış
gro.lo.la.mag	*NK*	kolola:ma:k
gra.myig.da	*NK*	kamığda:
gru.rug	*SY*	kuruğ
gra.yu	*SY*	kayu:
gog	*NK*	kök " heaven "
kyog	*Fr.*	kök " root "
go.gos	*NK*	kököz*
(*ti.kyin*) *kyol.miś*	*PN*	(tégin) külmiş
kyim	*SY, NK*	kim
kyon	*Fr.*	kün
kyor.sar	*SY*	körser
kyorg	*SY*	körk
kyi.śi.ke	*Fr.*	kişi:ke:
sa.grin.sar	*NK*	sakınsa:r
sa.gri[*nç*]	*SY*	sakınç
so.nar	*Fr.*	söner
sa.na.gu.lug.sus	*NK*	sanağuluğsuz
sas.sib	*NK*	"astonished," saşıb or şaşıb?

* This is certainly what was intended, but the MS. has *go. gro.* erasure *gos.*

yi.di	*Fr.*	**yéti:**
yog	*SY*	**yok**
yo.grug	*SY*	**yokuğ**
[*ya*].*gri.dın*	*SY*	**yağı:dın**
yag.ši	*PN*	**yaxşı:**
yu.ku.nur.myen	*NK*	**yükünürmen**
yil.diz	*SY*	**yıltız**
yi.ma	*SY*	**yime:** (or **yeme:?**)
yer	*NK*	**yér**

Tibetan *t* and *d* are very much alike, and in badly written texts almost indistinguishable; I have tried to transcribe what seems to be written, but I may well have made some mistakes. The convention of representing velar **k** by *gr* and the practice of using subscript *y* to indicate front vowels seem to show that the scribes were familiar with the manner in which the Brāhmī alphabet was used for writing Turkish, but the system of transcription is very rough and ready and extremely inconsistent. No attempt was made to use the scriptions for long vowels which were customarily used when Sanskrit words containing long vowels were written in Tibetan. Initial *p-* occurs only in two words and is unlikely to be significant phonetically. It may be more significant that **kök** " heaven " and **kögöz** " breast," which probably had an initial **g-** in pre-eighth century Türkish, are written with *g-* and **kök** " root," which certainly had an initial **k-**, is written with *ky-*, but this may be quite fortuitous. The only point of real interest in these scriptions is their confirmation that in Uyğur the vowels **o** and **ö** sometimes occurred in a second or subsequent syllable, but even in this the texts are not consistent, see the various scriptions for **tözön.**

(iv) THE TEXTS IN SOGDIAN AND UYĞUR SCRIPTS

The great majority of Uyğur and Uyğur-A texts which have survived are written in the script which is customarily called " Uyğur." This name is probably as anachronistic as that name when applied to the language, but in this case also no useful purpose would be served by suggesting

some other name. As was pointed out in Chapter II, the "Uyğur" language must have been well established in the steppes long before the Uyğur rose to prominence early in the eighth century, and it is almost certain that this script was devised for writing the language before that date. Most of these texts are Buddhist religious works, but there are also a good many legal and commercial documents written in this script, rather fewer medical and miscellaneous texts, a few Manichaean religious works and two or three fragments of Christian religious works. The most convenient list of published texts is in the list of abbreviations in A. Caferoğlu, *Uyğur Sözlüğü*, Istanbul, 1934. Almost the only substantial publications since that date have been further numbers of the *Türkische Türfan-texte*, which have now reached *X* and one or two texts in *Malov*, 1951.

There are also a few texts written in Sogdian script, and as the Uyğur script was derived from it, it will be better to start with the former. It was pointed out in the description of the Runic alphabet that it was to a large extent modelled on some form of the Aramaic alphabet which had been adapted for writing an Iranian language. The Sogdian is one such alphabet, but clearly not the one which served as the model for the Runic alphabet. Henning, *op. cit.*, traced the stages by which the Aramaic alphabet came to be used for writing Iranian and reached the conclusion (*op. cit.*, p. 30) that the transition from the use of the Aramaic alphabet to write Aramaic to its use to write an Iranian language, at first largely in cryptograms, probably took place in about the second century B.C. As time went on, fewer and fewer cryptograms were used (the tempo of change was slower in some languages than others), and by the time the Sogdians came into contact with Turks, and probably a good deal sooner, Sogdian had become an almost straightforward alphabetic script with very few cryptograms.

A tentative list of the phonetic values of the letters of the Aramaic alphabet has been set out above, and it was pointed out that in writing Iranian *tzaddi* was usually used

to represent ç and some letters were not used at all. Nevertheless in the schools the whole Aramaic alphabet was retained in its original order*, and no new letters were added. The reason for this conservatism was the fact that all the Aramaic letters occurred in one cryptogram or another, but in actually writing Sogdian the following letters (see Henning *op. cit.*, p. 59) were never used:—*daleth*†, *he* (which, however, survived without phonetic value, as a sort of determinative of words, even pure Sogdian words, in the feminine gender), *teth*, *ain* and *koph*.

Thus the effective Sogdian alphabet was reduced to seventeen letters. Even in the earliest form of the script which has survived, that of the "ancient letters" which can be firmly dated to the opening decades of the fourth century (see W. B. Henning, *The date of the Sogdian ancient letters*, B.S.O.A.S. XII, pp. 601 ff.) two of them, *gimel* and *cheth*, were barely distinguishable and two others, *zain* and *nun*, completely indistinguishable. In the variety of the script which was probably evolved in about A.D. 500 and has come down to us mainly in copies of religious books and so can be called "the Sutra script" this latter ambiguity was overcome by placing a diacritical point under *zain* to identify it,‡ but no attempt was made to identify *gimel* or *cheth*, which in this script had become indistinguishable, probably because by this time the sounds which they represented, *ǧ* and *x*, had become almost indistinguishable or even completely coalesced.

* The evidence for this statement is contained in a much garbled account of the Sogdian alphabet in a thirteenth century Persian work, see E. D. Ross and R. Gauthiot, *L'alphabet sogdien d'après un témoignage du 13e siècle*, Journal Asiatique, May-June, 1913, and E. Denison Ross, *Ta'rīkh-i Fakhru'd-dīn Mubārak Shāh*, Royal Asiatic Society (James Forlong Fund), 1927, page 44.

† This letter was not required because the voiced dental plosive *d*, which survived in the other Middle Iranian languages, had, except in one or two special contexts, become a spirant *ḏ* in Sogdian, and this sound, as explained below, was represented by *lamed*.

‡ It is important to distinguish the two different purposes for which diacritical points can be used. First they can be used to identify a letter the form of which is indistinguishable from that of another, as *zain* from *nun* here, or *bā* from *tā* and one or two other letters in the Cufic form of Arabic script. Secondly they can be used to indicate that a letter instead of having its usual phonetic value has a different but similar one, that is to differentiate two phonetic values of the same letter, say *b* and *p*, *t* and *ṭ* or *d* and *ḏ*. I have been careful to use "identify" in the first case and "differentiate" in the second.

In the Cursive script, which was probably evolved in the seventh century for ordinary day-to-day purposes, the practice grew up of identifying *zain* by disjoining it from the following letter, which made it possible to distinguish between initial and medial, but not final, *nun* and *zain*, and the diacritical point, which was no longer needed for identification purposes, began to be used to differentiate *zain* used to represent *z* from *zain* (with subscript dot) used to represent the voiced palatal sibilant *j*. Thus the effective Sogdian alphabet from about A.D. 500 contained sixteen, or at most seventeen, letters:—*aleph*, *beth*, *gimel-cheth*, *vau*, *zain* (not distinguishable from *nun* at the end of a word), differentiated *zain*, *jod*, *caph*, *lamed*, *mem*, *nun*, *samech*, *pe*, *tzaddi*, *resh*, *schin*, *tau*.

These letters had to be used to represent a wide range of sounds and a considerable amount of polyphony was inevitable. In giving the relevant phonetic values, I have followed R. Gauthiot, *Essai de grammaire sogdienne*, Paris, 1914-23, as corrected by Henning, *op. cit.*

VOWELS. Sogdian seems not to have had a glottal stop, words which did not begin with a consonant beginning with a smooth vocalic ingress; nevertheless the Sogdian script, unlike Runic, retained the obligatory use of *aleph* as an initial for all words beginning with vowels. The Sogdians also found it inconvenient to leave all short vowels unwritten, and began to represent many, but not all, of them by vowel letters. This in its turn blurred the distinction between short and long vowels, the former habitually left unrepresented, and the latter habitually written with vowel letters in Aramaic, and so a tendency developed to insert an *aleph* before the vowel letter representing a long vowel. Later still these extended representations of long vowels were sometimes also used to represent short vowels, and the spelling became quite chaotic. The full range of Sogdian vowel sounds seems to have been:—*a*, *a:*, *é:*, *i*, *i:*, *o:*, *u* and *u:*. In the table below the scriptions in brackets are, generally speaking, later than those not in

brackets. The actual scriptions, which varied according as the vowel was initial, medial or final, were as follows:—

	Initial	Medial	Final*
a	*aleph*	zero (*aleph*)	*aleph*
a:	*aleph aleph*	*aleph* (*aleph aleph*, zero)	
é:		*aleph jod* (*jod*)	
i	*aleph jod*	zero, *jod*	*jod*
i:		*aleph jod* (*jod*)	
o:		*aleph vau* (*vau*)	
u	*aleph vau*	zero, *vau* (*aleph vau*)	*vau*
u:		*aleph vau* (*vau*)	

It will be seen that most vowel letters and combinations of vowel letters represented more than one sound.

CONSONANTS. The consonantal structure of Sogdian seems to have been rather eccentric. In principle all plosives and affricates were unvoiced, but became voiced after nasals and in other rare consonantal clusters. On the other hand it was rich in fricatives and sibilants, both voiced and unvoiced. The old Iranian liquid *l* had become *r* in Sogdian as in other Middle North Iranian languages, but at any rate in the fourth century was still pronounced in loan words. On the other hand there must have been something peculiar about the pronunciation of the dental fricatives *ḏ* and *ṯ* which made the Sogdians feel that they could not appropriately be represented by *daleth*, and they represented them both, if they were still distinguishable (according to Henning, *op. cit.* p. 61, they may have almost coalesced) by *lamed*, which was no longer required to represent *l*, although still used in the " ancient letters " to represent that sound in Sanskrit loan words like *lekha* " letter " and *loka* " world."

The full range of consonantal sounds in Sogdian, excluding those occurring only in loan words, and putting

* All final vowels were short, but may not have been precisely *-a*, *-i* and *-u*. Gauthiot regarded final *he* as representing a short vowel, but, as stated above, Henning believes it to have been merely a determinative without phonetic value.

secondary sounds in brackets, can therefore be tabulated as follows:—

	Plosives		Fricatives		Nasals	Affricates		Sibilants		Semi-vowels
	u.	(v.)	u.	v.		u.	(v.)	u.	v.	
Labials	*p*	(*b*)	*f*	*v*	*m*	—	—	—	—	*w*
Dentals	*t*	(*d*)	*t̲*	*d̲*	*n*	—	—	*s*	*z*	—
Denti-palatals	—	—	—	—	—	*ʧ*	(*ʤ*)	—	—	—
Palatals	—	—	—	—	—	—	—	*ʃ*	*ʒ*	*y*
Post-palatals	*k*	(*g*)	—	—	—	—	—	—	—	—
Velars	—	—	*x*	*ǧ*	—	—	—	—	—	—

Liquid *r*.

It will be remembered that in the Aramaic alphabet, no doubt because the fricative sounds were of secondary origin, the same letters were used to represent plosives and homorganic fricatives (e.g. *beth b/v*), but that different letters were used to represent unvoiced plosives/fricatives (e.g. *pe p/f*) and voiced plosives/fricatives (e.g. *beth b/v*). In the Sogdian alphabet, however, presumably because it was the voiced sounds that were of secondary origin, the same letters were used for unvoiced and voiced plosives (e.g. *tau t/d*), but different letters for the homorganic fricatives (e.g. *lamed t̲/d̲*). There is one apparent exception; in the earlier texts *f* was represented by *pe*, as it was in Aramaic, but in the later ones by a slightly modified form of *beth*.

Thus the full range of phonetic values of the Sogdian alphabet, excluding sounds occurring only in loan words, and putting secondary sounds in brackets, was as follows:—

aleph	See table of representation of vowel sounds.
beth	*v*; later, in modified form, *f*
gimel-cheth	*ǧ*, *x*
vau	*w*; *o:*, *u*, *u:*
zain	*z*; when differentiated, *j*
jod	*y*; *é:*, *i*, *i:*
caph	*k*, (*g*)
lamed	*d̲*, *t̲*
mem	*m*
nun	*n*

samech	*s*
pe	*p*, (*b*); in the early period *f*
tzaddi	*ʒ*, (*c*)
resh	*r*
schin	*ʃ*
tau	*t*, (*d*).

Only very few Turkish texts are written in what can definitely be described as the Sogdian Sutra script, and these seem to be later than the earliest texts written in Uyğur script proper, and to have adopted the Uyğur graphic conventions. We can, therefore, leave them aside for the moment and pass on to the Uyğur script itself.

Henning (*op. cit.*, p. 56) points out that the vast majority of the Uyğur texts are written in a script which is close to the Cursive script, and in effect suggests that the two are really identical. But there is sufficient difference between them to justify the suggestion that the Uyğur really is a different alphabet which must have been deliberately devised to write Turkish either at some given point in time or in two or more successive stages, but the former is the more probable since there are no surviving specimens of a transitional script. It is unlikely that the precise circumstances in which this script was devised will ever be discovered. There is no reason, as there was in the case of Runic, to suppose that it was devised as an official alphabet. (Runic was the official alphabet both of the various Türkü Empires and of the first Uyğur Empire.) It can hardly have been originally a missionary alphabet as the Brāhmī, Tibetan and Manichaean Syriac alphabets were when used for writing Turkish, since it was used by Christians, Buddhists and Manichaeans alike. The remaining, and likeliest, explanation is that it was originally a commercial alphabet, devised by the Sogdian merchants for use in their Turkish correspondence. It was clearly modelled on the Sogdian Cursive script, and this gives a *terminus ante quem non* for the date at which it was devised. If Henning is right in saying that the Cursive script was first developed in the seventh century, then the Uyğur script cannot be

older than that, but it is unlikely that it is much younger. Sogdian merchants were certainly trading with Turks well before that time.

The points which distinguish the Uyğur from the Sogdian alphabet are:—

(1) the omission of the four Aramaic letters used only in cryptograms in Sogdian;

(2) the invention of a new letter, *resh* with a hook turning to the right attached beneath it, to represent the liquid **l**, which incidentally proves that by that time the use of *lamed* to represent *l* in loan words had been discontinued;

(3) the use of special devices mentioned below for representing initial **e-** and **ö/ü**.

All these devices occur in the Sogdian Sutra script in which a few Turkish texts are written, and it seems reasonable to suppose that these were the products of conservative professional scribes who felt more at home with their own familiar book script, but found it necessary to use these devices in writing in Turkish, in order to make what they had written intelligible.

The first point comes out clearly only in the few surviving examples of the alphabet written out as such. The earliest of these seems to be a scribble in the margin of a Manichaean text in Uyğur script transcribed rather imperfectly in A. von Le Coq, *Türkische Manichaica aus Chotscho* III, A.P.A.W., Berlin, 1922, p. 22. It is as follows:— *aleph*, *beth*, *gimel*, *vau*, *zain*, *cheth*, *jod*, *caph*, *lamed*, *mem*, *nun*, *samech*, *pe*, *tzaddi*, *resh*, *schin*, *tau*, hooked *resh*, and three letters in the final form used at the end of a word, *samech* (or *schin*?), *mem* and *gimel* (or *cheth*?); the last three letters can be restored from the later alphabet mentioned below. The manuscript, being Manichaean, is unlikely to be later than the ninth century. We are fortunate in having an eleventh century list of the letters of the Uyğur alphabet and a description of it in Kāşğarī; a translation of the relevant passage is attached as an appendix to this chapter. The alphabet contains eighteen letters, i.e. the seventeen

Aramaic/Iranian letters listed above and hooked *resh*. What the difference, if any, was between *gimel* and *cheth* in the Manichaean alphabet cannot be said in the absence of a facsimile, but in Kāşğarī *gimel* is the ordinary letter, transcribed in Arabic *xā*, and *cheth* is identified by two superscript dots and transcribed *qā*.

The Uyğur manuscripts which have survived cover a wide period of time, but few of them can be dated precisely even to within a century. A good many facsimiles of such manuscripts have been published in various editions of Uyğur texts, from which a good idea can be had of the way in which the script developed. The earliest, for example the Christian apocryphon in F. W. K. Müller, *Uigurica*, A.K.P.A.W., Berlin, 1908, and the Manichaean manuscripts edited with several facsimiles by A. von Le Coq may be as early as the eighth century and can hardly be later than the ninth, and some of the best Buddhist manuscripts may be equally old. This does not, of course, imply that the texts, and in particular the Buddhist texts, which they contain were composed at this date. It is uncertain when the Turks were converted to Buddhism, and equally uncertain in what script the earliest Turkish Buddhist texts were written. The former subject was recently discussed by Prof. von Gabain in her article, *Der Buddhismus in Zentralasien* in *Handbuch der Orientalistik* I, viii, *Religion*, 2, Leiden, 1961. She shows that a Buddhist *sūtra* was translated into Türkü in the second half of the sixth century, and presumably at that date written in the Runic script. But the later Türkü reverted to their original paganism, and it is unlikely that this text survived. The next firm date is A.D. 762 when the Uyğur kağan was converted to Manichaeism and destroyed a number of statues of the Buddha. By this time, therefore, Buddhism must have been well established among the Turks. This is confirmed by the fact that the Turkish Manichaean texts, the date of which will be discussed later, use a number of Turkish Buddhist technical terms. It is not, therefore, unreasonable to suggest that Buddhist missionaries were active in the

steppes among the T'ieh-lê and translating the scriptures into their language not later than the early part of the eighth century and very likely in the seventh century. The latest Uyğur manuscript, the well-known Turkish translation of the *Suvarṇaprabhāsasūtra*, is dated A.D. 1687, but this is a kind of fossil, a religious text re-copied long after Uyğur proper had ceased to be a spoken language. The next latest is probably document No. 22 in V. V. Radloff's *Uigurische Sprachdenkmäler*, Leningrad, 1928, which was written during the reign of Tuğluğ Témür (A.D. 1347-1363). Several other documents in that book are obviously later than the Mongolian invasion of the thirteenth century. Thus, if the manuscript of A.D. 1687 be left out of account, the script was in continuous use for everyday purposes for at least six centuries. A convenient, but rather summary, table of the Sogdian and Uyğur alphabets will be found in von Gabain *ATG*, page 17, followed by five facsimiles of manuscripts.

A study of the facsimiles will show how the script gradually deteriorated. In good early manuscripts it is reasonably easy to tell all the eighteen letters apart. *Samech* and *schin* have slightly different outlines; initial, and even medial, *aleph* and *nun* are just distinguishable, and *gimel-cheth*, although the two letters themselves are indistinguishable, is identified by two superscribed dots when it represents velar **k** (or **x**?). By the eleventh century *samech* and *schin* had become indistinguishable, but in carefully written manuscripts *schin* was identified by two subscribed dots. Similarly *nun*, no longer distinguishable from *aleph* or, as a final, *zain* was identified by one superscribed dot. Some other letters too were beginning to be difficult to distinguish. By the fourteenth century the only letters which were quite unmistakable were *caph*, *lamed*, *mem*, *pe* and hooked *resh*. In the absence of diacritical marks *aleph* and *nun*, and, in the final position, *zain* were indistinguishable. *Beth* and *jod*, and in badly written manuscripts even *tzaddi*, were barely distinguishable. Medial *gimel-cheth* was indistinguishable from two consecutive *aleph-nun*'s, and

the practice grew up of using two superscript dots, rather unmethodically, to identify this letter irrespective of its phonetic value. Medial or final *tau* was indistinguishable from *vau nun* unless the *nun* was dotted, so that for example only the context could decide whether a particular word was to be read as **et** " meat," **on** " ten " or **un** " flour." In the absence of diacritical marks, *samech* and *schin* were indistinguishable, and sometimes even difficult to distinguish from *gimel-cheth*. *Resh*, originally formed by two strokes crossing above the line and looking like a small V was sometimes only a rather bloated " tooth " hard to distinguish from *aleph-nun*. In addition, the phonetic fusion referred to below between *samech-schin* and *zain* and between *lamed* and *tau* had taken place and added to the difficulties.

With this rather discouraging introduction, we can now consider how the Uyğur alphabet was used to represent the Turkish sounds.

VOWELS. The convention of writing all words beginning with a vowel with an initial *aleph* was retained, but the practice of leaving short vowels unrepresented by vowel letters was almost completely abandoned. Short vowels were left unrepresented very sporadically, sometimes perhaps when they were super-short, as for example in **yeme:** " also " written *yme* (a scription similar to that in the Brāhmī script), but chiefly in Sogdian loan words which, like the Aramaic " cryptograms " in Sogdian, were taken over in their original spelling. The practice of inserting an *aleph* before medial long vowels, which was no longer used in the Sogdian script for that purpose only at the time when the Uyğur script was developed, was abandoned, and thus no method existed of distinguishing long vowels from short ones. A method was devised of distinguishing initial **a-** from **e-** by writing the former *aleph aleph** and the latter *aleph*; but medial and final **-a-/-a:** and **-e-/-e:** were both written *aleph*. No attempt was made

*Some cases occur of **a-** being written *aleph;* the reasons for this *scriptio defectiva* are obscure.

to distinguish between ı and i. As regards the rounded vowels, in order to distinguish them from o/u, ö/ü were written *vau jod*, but only in the first syllable. This is the exact converse of the practice in the Brāhmī script; the device of using *jod* in a non-phonetic manner to distinguish the quality of the adjacent vowel letter is unlikely to have been invented twice quite independently, but there is no means of ascertaining in which script, Brāhmī or Sogdian, it was first introduced. The vowels were therefore written in the following manner, irrespective of their length:—

	Initial	Medial and final
a	*aleph aleph*	*aleph*
e	*aleph*	
é	*aleph jod*	*jod*
ı		
i		
o, u	*aleph vau*	*vau*
ö, ü	*aleph vau jod*	(1st syllable) *vau jod*; (elsewhere) *vau*.

It will be seen that all the vowel letters were polyphonic.

CONSONANTS. Reference has already been made to the invention of a new letter, hooked *resh*, to represent l. The non-Iranian nasals, ŋ and, where it survived, ñ, were represented by two letters, *nun kaph* and *nun jod* respectively. The use of diacritical marks at various periods has been described above. Thus, subject to the notes below, the seventeen (or eighteen) letters of the Uyğur alphabet, when used to write Turkish words (as opposed to loan words containing non-Turkish sounds), had the following phonetic values:—

aleph	see table of representation of vowel sounds.[1]
beth	v (?f).
gimel-cheth	ğ, velar k, ?x.
vau	o/u; ö/ü.

zain	z.[2,3]
jod	y; see table of representation of vowel sounds.
caph	g, post-palatal k.
lamed	d, d̲.[4]
mem	m.
nun	n[3] (*nun*, *caph* ŋ; *nun jod* ñ).
samech	s.[3]
pe	b, p.
tzaddi	ç (? c).
resh	r.
hooked *resh*	l.
schin	ş.[3]
tau	t, d.[4]

NOTES:

1. See above for the methods of identifying *aleph* and *nun*.

2. As in Sogdian, *zain* was not joined to the following letter.

3. In the earlier manuscripts *zain* represented z (and in loan words *j*, the latter usually differentiated, as in Sogdian, by one subscribed dot), *samech* s and *schin*, identified by two subscribed dots when it ceased to be distinguishable, ş. At some date, perhaps even as early as the eleventh century, the practice grew up of using all three letters indifferently to represent all three (four) sounds, the single and double dots sometimes being placed under whichever letter was used in order to differentiate it when it represented j or ş. It is significant that in the alphabet in Kāşğarī, see the appendix to this chapter, *zain* and *samech* are in their right places in the list of letters shown separately but have changed places in the specimen of continuous script.

4. It will be remembered that in Sogdian *d* was a secondary sound occurring only in the combination *nd* and in other rare consonantal clusters, where it was represented by *tau*, which normally represented *t*. The same practice was followed in the early Uyğur manuscripts,

where **t** was represented by *tau* and **-nd-** and **-nt-** (a sound not existing in Sogdian) were represented by *nun tau*. Apparently *tau* also represented **d** in some other contexts, e.g. **-rd-** and **-rt-** both represented by *resh tau*. On the other hand both **ḏ** and intervocalic **-d-**, a sound not existing in Sogdian, were represented by *lamed*. At a later period, probably the same as that mentioned in the preceding note, both *lamed* and *tau* came to be used indifferently to represent **d, ḏ** and **t** in all contexts, the scribe choosing whichever of the two he found it most convenient to write on each occasion. This is a point on which Kāşğarī's evidence would have been valuable, but he transcribed *lamed ḏā* and *tau tā* and did not mention **d**.

It should be emphasized that even in the best later manuscripts the use of diacritical points is unmethodical and sporadic, and in the late commercial and legal documents almost none are found. As in the Runic script, double consonants, when they occurred, were normally written with a single letter; but in some manuscripts suffixes like **-lığ/-lig** were disjoined from the word to which they were attached, and in words like **allığ** " crafty " both **l**'s were written.

It is obvious that an alphabet as polyphonic as this is practically useless as a guide to the phonetic structure of the languages which it was used to write. It is superior to Runic only on one point, it has distinct letters for **b** (*pe*, also **p**) and **v** (*beth*). In all other respects it is greatly inferior to the other alphabets enumerated in this chapter. It is possible to reconstruct the phonetic structure of Uyğur with some confidence only because there are no very large phonetic differences between Türkü and Xākānī, and Uyğur lies more or less squarely between the two.

(v) THE TEXTS IN MANICHAEAN SYRIAC SCRIPT

The most convenient and succinct account of Mani and Manichaeism will be found in A. Christensen, *L'Iran sous les Sassanides*, second edition, Copenhagen, 1944, Chapter IV. Mani was born in Babylonia in A.D. 216 or 217 of

parents who were Iranian by origin but spoke Syriac as their mother tongue, and it seems probable that his early works, which have not survived, were in Syriac, but his later works seem to have been composed in some Iranian language, probably Middle Persian rather than Parthian, although texts in both these languages have survived. It is therefore easy to see why the "official" script of Manichaeism was a rather ornate form of the Syriac Estrangelo script adapted for writing Iranian. This alphabet was descended from the original Semitic alphabet of twenty-two letters described above, but by a different line of descent from that of the Aramaic alphabet, although the actual Syriac and Aramaic languages were very similar. The letters of the two alphabets therefore look quite different. It seems that when the Syriac alphabet was adapted for writing Iranian languages, very possibly by Mani himself, in the third century A.D., rather different methods were used from those which had been used some centuries earlier when the Aramaic alphabet was adapted for this purpose. The subject is discussed in detail in Henning, *op. cit.* p. 74. In the earliest period it seems that the Syriac alphabet as such was used in spite of its polyphonic character; later some attempt was made to abate this polyphony by putting dots over some letters to differentiate them and by using other letters in slightly different shapes to represent different sounds. The alphabet was adapted to transcribe at least three different Iranian languages, Middle Persian, Parthian and Sogdian, with slightly different phonetic structures, and the differentiated and altered letters varied according to the language being transcribed. A convenient table of the form of letters used to write Iranian languages will be found in F. W. K. Müller, *Handschrift-reste in Estrangelo-schrift aus Turfan, Chinesisch-Turkistan*, II, A.K.P.A.W., 1904, page 5. When the alphabet was adapted by the Manichaean missionaries in Central Asia for writing Turkish it is not surprising that, since those missionaries were no doubt Sogdians, it was primarily the variety used for writing Sogdian that was taken as a model.

There were only two Turkish consonantal sounds for which no Iranian equivalent existed, velar **k** and **ğ**; for the first both *caph* and *koph*, which had been used indiscriminately to represent Iranian post-palatal *k*, were adopted, differentiated by two superscribed dots; for the second a slightly altered form of *gimel* was devised. There were also, of course, some Turkish vowel sounds which did not exist in Iranian, and to represent them the same devices were employed that were employed in the Uyğur script. This proves, if proof were needed, that the Uyğur script was in use before the Manichaean missionaries arrived among the Turks.

The same combinations of *aleph*, *vau* and *jod* as those used in the Uyğur script were used to represent the Turkish vowel sounds with one minor alteration. Initial **é-**, **ı-** and **i-** were usually, but not always, represented by *ain jod* and not *aleph jod*. The Manichaean Syriac letters used to write Turkish will be found in the combined table in von Gabain *ATG*, p. 17, accompanied by a facsimile of a text in this script. A comparison between this table and that in F. W. K. Müller, *op. cit.*, shows that the only letters in the Iranian list which do not appear in the Turkish list are *cheth*, used for *x*, the singly dotted *pe* used for *f* and the altered form of *tzaddi* used for *c*.

The letters were used in the following way to represent Turkish sounds:—

aleph	as in the Uyğur script.
beth	**b.**
beth	with two superscribed dots **v.**
gimel	**g.**
altered *gimel*	**ğ.**
daleth	**d, <u>d</u>** (?), very rarely used.
(*he* used only to fill empty spaces at the end of lines.)	
vau	**o, u, ö, ü.**
zain	**z.**
teth	**t.**
jod	as in the Uyğur script.
caph	post-palatal **k.**

caph	with one superscribed dot **x**, very rarely used.
caph	with two superscribed dots velar **k**.
lamed	**d**, **d̲**.
altered *lamed*	**l**.
mem	**m**.
nun	**n** (*nun caph* **ŋ**; *nun jod* **ñ**.)
samech	**s**.
ain	only in *ain jod* **é-**, **ı-**, **i-**.
pe	**p**.
tzaddi	**ç** (and **c**?).
koph	post-palatal **k**.
koph	with one superscribed dot **x**, very rarely used.
koph	with two superscribed dots velar **k**.
resh	**r**.
schin	**ş**.
tau	**t**, seldom used.

It will be seen that, so far as consonants are concerned, the full range of Turkish sounds was covered, and that except for *lamed*, and perhaps *daleth*, no letter represented more than one sound. Unfortunately some of the scribes were rather careless or unskilful and the dots were often omitted. As in the Runic and Uyğur scripts double consonants, when they occurred, were normally written with a single letter.

Most Manichaean texts are written in this script, one or two in the Sogdian and the rest in the Uyğur script. The longest continuous text is the almost complete manuscript of the so-called *Chuastuanift* (properly *Xwastwānēft*) acquired by Sir Aurel Stein in Tunhuang and published by A. von le Coq with a facsimile and translation in J.R.A.S., 1911, pages 277 ff. Nearly all the other Manichaean Turkish texts were published in A. von le Coq, *Türkische Manichaica ans Chotscho*, I to III, A.K.P.A.W., Berlin, 1912, 1919, 1922, and in *T.T.*, II, III and IX, Berlin, S.P.A.W., 1929 and 1930, A.D.A.W., 1958. The last text, though very fragmentary, is of exceptional importance, since it is a bilingual, similar to the Sanskrit-Turkish bilinguals in Brāhmī script,

but in this case the second language is "Tokharian B," the Indo-European language spoken in Kucha.

Some of the texts contain very unusual spellings; for example in the first text in *TMC* I **s** is consistently substituted for **ş**, and in the large but fragmentary manuscript of which parts were published in *TT* III and IX there is a series of substitutions, **p** for **b** and so on, which are discussed in the introductions to those volumes. Prof. von Gabain's theory is that these represent the peculiarities of some Turkish dialect; this is of course possible, but I am inclined to think it more probable that they indicate that the scribes were not Turks and could not spell Turkish properly. This is supported by the fact that similar mistakes occur in the spelling of the "Tokharian B." Some of the sound changes concerned, for example the representation of voiced **b** by unvoiced **p**, are the kind of mispronunciations of which a Sogdian might be guilty, and the scribes may well have been Sogdians.

Prof. von Gabain, in *TT* IX, pages 6 ff., discusses at some length the date at which some Turks were converted to Manichaeism and the date of the manuscript published in that volume. The date of the official conversion of the Uyğur *kağan* to Manichaeism was A.D. 762, but this was rather the culmination than the beginning of Manichaean missionary enterprise among the Turks, and she is inclined to date the first conversions to a rather earlier date than this, but probably not earlier than A.D. 700. The actual manuscript she is inclined, taking into account factors like the quality of the paper, to date to the middle of the tenth century, but the text which it contains must have been composed much earlier than this. A good deal turns on the dialects in which the texts are written.

Prof. von Gabain has pointed out in *ATG* that, quite apart from the eccentricities of spelling already mentioned, which probably do not reflect differences of dialect, three different dialects can be distinguished. Some, perhaps most, texts, including those in *TT* III and IX, are in standard Uyğur; others, probably rather fewer, are in Uyğur-A. A

good many texts, including the *Chuastuanift*, are written in a third dialect which is remarkable for the fact that the old palatal nasal ñ has survived in it, spelt sometimes *n*, sometimes *ny* and sometimes *yn*. This dialect has other phonetic and grammatical peculiarities which it shares with the Orkhon inscriptions and manuscripts in Runic script. There can in fact be little doubt that it is a variety of Türkü and can reasonably be called "Manichaean Türkü." It is not always possible to decide in which of the three dialects a short text is, but the position seems to be that most of the texts in Manichaean Syriac script are in Manichaean Türkü (those in *TT* III and IX are the most important exception) and the texts in Uyğur script are more often in Uyğur or Uyğur-A.

It seems reasonable to suggest that there is a good historical reason for this phenomenon, probably that the Manichaean missionaries began to translate their scriptures into Turkish before the collapse of the Türkü Empire in A.D. 742, and that the texts in Manichaean Türkü were translated before that date. Which came next perhaps depends on the question whether the language of the Uyğur themselves was "Uyğur" or "Uyğur-A," but whichever it was it seems reasonable to assume that some translations were made at the command of the Manichaean Uyğur *kağan* in his particular dialect, while the rest were made for the use of converts belonging to those tribes who spoke the other dialect. On the whole it seems more probable that the language of the Uyğur themselves was actually Uyğur-A, but this cannot be more than a conjecture.

The fact that the texts in Manichaean Türkü are mainly written in Manichaean Syriac script supports the suggestion that these are the earliest Manichaean texts, since it is very probable that the mssionaries used their own script for their earliest translations and only later adopted the Sogdian and Uyğur scripts.

Considered as a guide to the phonetic reconstruction of the dialects for which it was used, the Manichaean script,

provided that it is carefully written with the necessary dots, is superior to Runic in distinguishing between **b** and **v**, slightly inferior to it in distinguishing between back and front vowels and **é** and **i**, and equal to it in every other respect. It is in every respect superior to the Sogdian and Uyğur scripts. But the value of some texts is diminished by the omission of some dots and by eccentric spellings.

(vi) TEXTS IN ARABIC SCRIPT

The only text in Arabic script which provides important and accurate material for the reconstruction of the phonetic structure of pre-eighth century Turkish is Kāşğarī's *Dīwān*, since it is the only one which systematically distinguishes between long and short vowels. Kāşğarī was a competent philologist, profoundly interested in the Turkish language, and he made every effort to write Turkish words as exactly as possible, making full use of the vowel signs (*ḥarakāt*) and vowel letters. Even so his work is not a perfect instrument for this purpose, since, while the words constituting the main entries of the dictionary are spelt with great care, the words in phrases and quotations illustrating these entries are spelt less carefully, with the result that some words are vocalized differently in the main headings and in the illustrative matter, even when they are in the same unsuffixed form. This raises some doubt whether the vocalizations of words carrying declensional suffixes, when they differ from those of the unsuffixed words, genuinely indicate a change of vocalization, and specifically an alteration in the length of vowels, caused by the attachment of suffixes, or are merely the result of carelessness. Similarly there is a doubt whether the monosyllabic verbs, which are seldom recorded in the unsuffixed form (the second person singular of the imperative) and are noted in suffixed forms with short vowels only, even when such vowels are followed by a consonant (for example a voiced plosive) which in a noun normally follows a long vowel, really had a short vowel in the unsuffixed form, or whether they had in it a long vowel

which was shortened by the attachment of a conjugational suffix. Nevertheless, in spite of these doubts, Kāşğarī is an invaluable guide to the phonetic structure of Xākānī, and therefore an indirect source of information on the phonetic structure of earlier stages of the language.

The other early texts in Arabic script, like the *Kutadğu: Bilig*, although they often provide useful information about consonantal values, notably by the use of triply dotted *fā* to denote **v**, so far as the vowels are concerned were written in an orthography which was more designed to identify the vowels than to indicate their length. It is commonly thought that this wide use of vowel letters to denote short vowels was a legacy from Uyğur orthography, but it may equally well have been a deliberate policy of the scribes, who made little or no use of *ḥarakāt* and assumed that the reader would be able to supply the right short vowels when reading an Arabic text, but would be at a loss when reading a Turkish one so written. The practice in Central Asia differed greatly from that in Anatolia, where the *ḥarakāt* were widely used, at any rate till the middle of the fifteenth century. It is possible that a careful study of these early Anatolian manuscripts might throw some light on the question whether there was a distinction between long and short vowels in Old Osmanli. The question has never, so far as I know, been explored, because it has always been supposed that, as there was no such distinction in later stages of the language, there cannot have been one in the earlier stages, but this is really a *non sequitur*, particularly since long vowels have survived until the present day in Türkmen, which is a closely related language.

The Arabic alphabet is so well known that it is unnecessary to discuss it in detail. So far as the vowels are concerned, it could distinguish between short and long vowels; its great defect was that it only had scriptions for three of each, although there is little doubt that even in Arabic itself there were more than three, and it is certain that in Persian, Turkish and other languages for which it was used there was a much wider range of vowel sounds.

Thus when it was used for representing Turkish all the vowel signs and vowel letters were polyphonic:—

fatḥa	**a, e**	*alif*	**a:, e:**
kasra	**é, ı, i**	*yā*	**é:, ı:, i:**
ḍamma	**o, u, ö, ü**	*wāw*	**o:, u:, ö:, ü:**

So far as consonants are concerned, the actual letters used by Kāşğarī for representing Turkish sounds were:—

bā	**b, p**	*fā*	**f** (rare; not a Turkish sound?)
tā	**t**		
cīm	**ç** (occasionally **c**)	*fā*,	triply dotted, **v**
xā	**x** (rare)	*qāf*	velar **k**
dāl	**d**	*kāf*	**g**, post-palatal **k**
ḏāl	**d̲**	*lām*	**l**
rā	**r**	*mīm*	**m**
zayn	**z**	*nūn*	**n**
zayn, triply dotted,	**j** (not a Turkish sound?)	*nūn kāf*	**ŋ**
		wāw	**w** (rare and dialectic)
sīn	**s**	*hā*	**h** (rare, not a real Turkish sound)
şīn	**ş**	*yā*	**y**
ğayn	**ğ**		

Kāşğarī (*I*, 8) mentioned the existence of **p**, calling it *al-bā'u'l-ṣulba*, and elsewhere very occasionally referred to its existence in particular words, but never used the Persian letter triply dotted *bā*. He said that **c** was a rare sound, but did not mention the circumstances in which it occurred. The words in his dictionary were arranged in sections in such a way that there is seldom, if ever, doubt whether *nūn kāf* was intended to represent **ŋ** or **ng**. He said that **h** was not really a Turkish sound, but occurred in one or two words, mainly loan words. It will be seen that only *bā*, *cīm* and *kāf* were polyphonic, and so far as the consonants were concerned the alphabet was therefore superior to almost all the others for writing Turkish.

(2) TURKISH WORDS AND NAMES QUOTED IN FOREIGN AUTHORITIES

These authorities fall into two groups, eastern and western.

(i) THE EASTERN AUTHORITIES

Turkish names, words and very occasionally phrases occur in the Chinese authorities, mainly the dynastic histories, in Tibetan texts and documents, and in Saka documents.

(a) THE CHINESE AUTHORITIES

The difficulties of interpreting Chinese scriptions of Turkish and other foreign names, words and phrases were discussed at length in Chapter I. Until a good deal more progress has been made in publishing the original records and in working out the phonetic values of the characters used in these scriptions in each of them, it is doubtful whether much information of value for the reconstruction of the phonetic structure of pre-eighth century Turkish will be obtained from them. It would in any event be idle to suppose that these scriptions will throw much light on the niceties of Turkish pronunciation, but they may be of some help where there was a fundamental difference between eighth century Turkish and earlier stages of the language, for example in establishing in what words there was originally an initial **d̲-** or **ñ-** which had become **y-** in eighth century Turkish.

(b) TIBETAN TEXTS AND DOCUMENTS

Most of the early Tibetan material which contains references to the Turks was collected by Prof. F. W. Thomas in his *Tibetan Texts and Documents* in the Royal Asiatic Society's Oriental Translation Fund series, and in *Documents de Touen-houang rélatifs à l'histoire de Tibet*, Paris, 1940-1946, in which he collaborated with J. Bacot and Ch. Toussaint; but considering the very rudimentary methods of transcription employed even in the Turkish texts in

Tibetan transcription discussed in (1) (iii) of this chapter there is not much prospect that these documents, which are after all no older than the Runic texts, will give us much help. Moreover, it should be noted that sometimes what the Tibetan documents give us is not a direct phonetic representation of a Turkish word or name in Tibetan script, but a Tibetan transcription of a Chinese transcription of that word or name.

(c) SAKA DOCUMENTS

The Saka documents containing Turkish names and words have all been edited by Sir Harold Bailey, some with translations. A bibliography will be found in his article *Languages of the Saka*, quoted in (1) (ii) of this chapter. They include parts of a Turkish-Saka vocabulary. These documents are not as old as some of the surviving Turkish texts and, in view of the roughness of the transcription and the doubts regarding the phonetic value of some of the letters used, it is doubtful whether they can supply any very useful information on the niceties of Turkish phonetics. They contain of course a good deal of material interesting from other points of view.

(ii) THE WESTERN AUTHORITIES

The main western authorities which are old enough to be of interest for the particular subject under discussion are the Byzantine histories, chronicles and the like. We are so fortunate as to have in Gy. Moravcsik, *Byzantinoturcica*, Budapest, 1943; 2nd edition, Berlin, 1958, a most painstaking and comprehensive collection of Turkish names, words and phrases contained in these authorities with ample references to modern commentators. The greater part of them are of course much too late to present points of interest, but one author at least, Priscus, is as early as the fifth century and others are not much later, and these early texts contain some words of great interest. Reference has already been made in Chapter II to the importance of names like Ougouroi for throwing light on the early

history of l/r Turkish. Generally speaking, as in the case of the Chinese authorities, the methods of transcription are too rough to throw much light on the niceties of Turkish phonetics, but they contain some evidence of importance concerning major sound changes. The most important single piece of evidence is the statement of Menander Protector, a sixth century author, that the Türkü called "a funeral feast" *ḏogia* (there is no doubt that at this period *delta* represented the voiced dental fricative and not a plosive). This is unquestionably **ḏoğ*, an earlier form of **yo:ğ**, a word noted in the same sense in *KT* and *BK*. It has also been suggested that the earliest recorded name of the Ural River is Turkish. This theory seems to have become generally accepted, the most recent reference to the subject being in L. Ligeti's review of the second edition of Moravcsik's book in *Acta Orientalia Hungarica* X, but it seems to me open to the gravest doubt. The name first appears in Ptolemy (second century A.D.) as *ḏaiks*, of which the *-s* is presumably the Greek nominative suffix; in Menander Protector it is *ḏaix*, and in Constantine Porphyrogennetos (tenth century) *gee:x*. It has been suggested that this represents a Turkish word **ḏayık* >*yayık* "winding, bending" from the verb **ḏay-* >*yay-*. I have an uncomfortable feeling that this is merely an imaginative false etymology. I have traced the theory back to J. Marquart, *Die Chronologie der alttürkischen inschriften*, Leipzig, 1898, p. 76, where there is a reference to a work by Tomaschek which I have not been able to consult. We can accept the fact that in the second century A.D. the local name of the river was *ḏaik* or *ḏayik*, but it is now generally accepted that at that date the inhabitants of that part of the world were Sarmatians, and it is *prima facie* improbable (see Chapter I) that at that time any Turks were sufficiently well established in the vicinity of the Ural river to impose their own names on natural features. It is much more probable that the name was a Sarmatian one. It is quite true that there is evidence in the Arab geographers that the Ural river was later, say in the tenth century, called *yayik*

and this agrees with the Byzantine form *gee:x*, but I cannot find any evidence that the Turkish word *yayık*, which primarily means " fickle," for example of fortune, was ever used by the Turks as the name of a river, or specifically of the Ural river. Moreover there is no Turkish verb *yay-* which could at an earlier date have been ***ḏay-**. In the eleventh century or later there were two verbs **yay-**, which in the eighth century were **ya:ḏ-** and **ya:ñ-** respectively; neither of these could earlier have been ***ḏay-**; if they had originally had an initial **ḏ-** the forms would have been **ḏa:ḏ-** and **ḏa:ñ-**. The most, I suggest, that can be claimed is that if the Ural river was called *yayik* instead of *ḏayik* when the local population became Turkish it was because the Turks could no longer pronounce initial *ḏ-* and made it *y-* even in foreign words.

The other important block of western evidence is the remains of Proto-Bulgar discussed in O. Pritsak, *Die Bulgarische Fürstenliste und die Sprache der Proto-bulgaren*, Wiesbaden, 1955, and J. Benzing, *Das Hunnische, Donaubolgarische und Volgabolgarische* in *Philologiae Turcicae Fundamenta*, Wiesbaden, 1959. Both these, and particularly the second, have been subjected to a good deal of criticism, some of which is clearly justified, but there does not seem to be any real doubt that the Proto-Bulgar word *ḏilom* is ***ḏıla:n** > **yıla:n** " snake," and that a few other items of interest can be extracted from this material.

(3) EARLY TURKISH LOAN WORDS IN FOREIGN LANGUAGES

The language from which most evidence of this kind can be obtained is Mongolian. In *The earliest Turkish loan words in Mongolian* and *The Turkish elements in 14th Century Mongolian* I discussed the earliest layer of Turkish loan words in that language and I shall return to the subject in Chapter XI. These words are particularly interesting because they were borrowed from an early **l/r** Turkish language, probably Tavğaç, the phonetic peculiarities of which I shall discuss in that chapter. The evidence avail-

able, chiefly that contained in the Mongolian *Hua-I i-yü* of A.D. 1389 and the thirteenth century *Secret History*, which has come down to us in a transcription in Chinese phonetic characters made towards the end of the fourteenth century, is not in a form which gives us much guidance on the finer points of phonetics.

The really interesting feature of these loan words is that they were taken, probably in the fifth or sixth century, from a language which, although it had suffered certain phonetic changes, was in other respects extremely archaic, and in particular still contained initial **d̲-** and **ñ-** in words in which these sounds have both become **y-** by the eighth century at any rate in Türkü and the other standard Turkish languages. In this respect the language was in the same stage of phonetic development as that for which we have evidence in the Chinese and western authorities already mentioned. The words concerned are not numerous, but there are enough to make the existence of these initials certain. What they do not enable us to do is to estimate what proportion of the words which began with **y-** in the eighth century had earlier begun with these initials, since it does not seem to be possible to identify with certainty any words beginning with an original **y-** which entered Mongolian at this early period. It should, however, be said that in other positions **ñ** always seems to have been rare, but **d̲** and **y** relatively common sounds.

The only other important language which is known to contain a number of very early Turkish loan words is Magyar. For many years Hungarian scholars have been studying the Turkish (and other) loan words in Magyar and they have made many important discoveries. Unfortunately their discoveries have been for the most part published in their own language, with which regrettably few scholars outside Hungary are acquainted. However some works in other languages have been published. By far the most important is Z. Gombocz, *Die bulgarisch-türkischen Lehnwörter in der ungarischen Sprache*, published by the Société Finno-Ougrienne, Helsingfors, 1912. Many

more important discoveries have been made since that date. The most recent publication on the subject is Prof. L. Ligeti's *A propos des éléments "altaiques" de la langue hongroise*, Acta Linguistica Hungarica, XI, 1-2, 1961, which conveys the welcome news that plans are on foot to publish a French translation of a book of major importance by Prof. G. Bárczi.

It is generally agreed that the Magyars in the course of their wanderings from their original home in the neighbourhood of the Urals to Hungary were at different times in contact with various Turkish tribes speaking languages with different phonetic characteristics. It is also agreed that the earliest layer of loan words contains words borrowed from an **l/r** language of the same type (though of course not the same language) as that which provided the earliest layer of loan words in Mongolian. The date at which, and the tribe from which, these words were borrowed is still a matter for discussion. We know that the Byzantines were already in the fifth century in contact with Turks speaking an **l/r** language somewhere in the neighbourhood of the Black Sea, but it is doubtful whether the Magyars had left their original home as early as this. It seems to be agreed that they had already reached the Black Sea area by the ninth century and were in contact with Bulgars on the Kuban River, but these words may have been borrowed before that date. At a later date the Magyars seem to have been in contact with some Turkish tribe which had made the sound change initial **y-** > **c-**. Complicated questions both of Magyar and of Turkish phonetics are involved, but it seems easier to derive Magyar *gyöngy* "pearl" from **yinçü:** "pearl" (Türkü) through an intermediate form **cinçü:** than from a hypothetical earlier form ***dinçü:**.

There is sufficient evidence to prove conclusively that the first layer of loan words was borrowed from a language which still preserved initial ***d-** and ***ñ-**. So far as ***d-** is concerned, the clearest case is *dió* "nut" < ***dağak** > **yağak** (Uyğur). There is also good reason for supposing

that *disznó* " pig " goes back to a Turkish word ***d̲isna:ğ.** No such word has survived in the standard Turkish languages, where " pig " is either **lağzın** or **toŋuz (doŋuz),** both occurring in Türkü, but in Chuvash " pig " is **sısna,** which can be shown to go back to some such form as **d̲ısna:ğ.**

There are rather more words which preserve the original ***ñ-.** The clearest case is *nyár* " summer " < ***ña:z** > **ya:z** (Türkü). Ligeti, *op. cit.*, mentions one or two others. I suggested to him that *nyír-* " to shave, shear " could be connected with **yar-** (? **ya:r-**) " to split, cleave " (Xākānī) but he felt that, although no satisfactory Finno-Ugrian etymology for this word had been found, the semantic connection was too remote. He also informed me that a satisfactory Finno-Ugrian etymology existed for *nyúz-* " to skin," which excluded the possibility of a connection with **yüz-** (**yü:z-**) " to skin " (Xākānī). The connection would in any case have been difficult to establish since a form **ñü:z-** could not have existed in an **l/r** language of the type from which the other words were borrowed.

This is of course no more than an outline of the kind of information regarding early Turkish phonetics which can be gleaned from Magyar; the subject is too large and complicated to be dealt with here.

(4) MODERN TURKISH LANGUAGES

The evidential value of the modern Turkish languages arises from the fact that they have been written down in alphabets which, even though some of the letters are still polyphonic, give a much more precise idea of the phonetic structure of the languages concerned than any of the ancient alphabets enumerated above. The disadvantage of using them as evidence arises from the mere fact that they are modern, and that if they differ phonetically from earlier languages it is sometimes difficult to decide whether this is because they have undergone phonetic changes comparatively recently or because they retain an archaic

phonetic structure which the early languages concerned had lost before the date from which we first have specimens of them. But there is one case in which it seems clear that some modern languages do retain very archaic phonetic traits. Several modern languages have both voiced and unvoiced initials in the labial, dental and post-palatal series, while the early languages have only one or the other. It is quite usual when there are two kinds of sounds like this for the rarer to be assimilated to the commoner, but there can be no logical reason why, if originally all the initial sounds in a particular series were voiced (or unvoiced) the initial sound should become unvoiced (or voiced) in some words but not in others.

The modern languages concerned fall into two classes, the literary languages, the spelling of which has been normalized, and the spoken dialects of those languages, which careful recording by qualified investigators often shows to have had a much more variegated phonetic structure. Turkish dialectology is an enormous subject, the scientific study of which has so far not proceeded very far. Russian scholars first embarked on it systematically in the second half of the nineteenth century and the great series edited by V. V. Radloff, *Obraztsy Narodnoy Literatury Tyurkskikh Plemēn* (" Specimens of the popular literature of the Turkish tribes ") was the first substantial venture in this field. The vocabularies of these texts were to a large extent, but not completely, incorporated in his *Opyt*. Dr. Jarring and other scholars have collected and published a good deal of material from Sinkiang. Turkish scholars have collected and published a great deal of material regarding the dialects of Anatolia and Rumelia, and the vocabulary material emerging from it has been published in *Söz Derleme Dergisi*, 1st edition (5 volumes and index), İstanbul, 1939 ff.; 2nd edition now in course of preparation. The present position and future tasks of Turkish dialectology in the Soviet Union are discussed in a series of monographs in *Voprosy Dialektologii Tyurkskikh Yazykov*, Baku, 1960. Prof. N. A. Baskakov's *Tyurkskie Yazyki*, Moscow, 1960,

includes an admirable short study of the modern Turkish languages and their dialects.

The evidential value of the modern Turkish languages varies a great deal. In some of them the phonetic structure has been almost completely standardized and normalized with a comparatively limited range of sounds. For example, the difference between long and short vowels has been almost completely eliminated; it still survives however in Türkmen and perhaps one or two other languages, and in other languages like Chuvash and Yakut the old long vowels, or some of them, have evolved into diphthongs by a process which does not yet seem to be entirely understood.

Again in the field of consonants there has been a great deal of levelling. For example, most languages have only one initial labial plosive, in some **b-** and in others **p-**. Even in languages which are closely related the sound may be a different one; in Khakas all the words beginning with **b-** seem to be loan words, and all the native Turkish words begin with **p-**, while in Tuvan the overwhelming majority of words beginning with **p-**, perhaps all of them, are loan words and all the native Turkish words begin with **b-**. This is at any rate true of these languages as recorded in the official dictionaries, but it is clear from Baskakov *op. cit.* that the phonetic structure of some of the spoken dialects in the group is quite different; in some Northern Mountain Altai dialects there are no initial labial plosives, the sound being replaced by **m-**.

It is obvious that languages which have been subjected to such thorough-going levelling cannot be expected to provide reliable evidence of the finer points of Turkish phonetic structure. I did however discover when I was collecting the material for *The initial labial sounds in the Turkish languages* that some modern languages, particularly those descended from Old Oğuz, have retained a more variegated phonetic structure. In Osmanli/Republican Turkish, for example, most Turkish words with an initial labial plosive begin with **b-**, but quite an appreciable minority begin with **p-**, and words which begin with **p-** in

this language also begin with **p-** in one or two others, even though in them the normal initial labial plosive is **b-**, and although initial **p-** seems to be unknown in ancient languages like Türkü, Uyğur and Xākānī. Similarly I am beginning to collect evidence that although the normal initial dental plosive is **t-**, an appreciable number of words in some languages begin with **d-**, and that although the normal initial post-palatal plosive is **k-**, an appreciable number of words in some languages begin with **g-**. There seems therefore to be a good hope that a close study of the phonetic structure of some modern languages and dialects may disclose facts regarding the phonetic structure of pre-eighth century Turkish which cannot be obtained from any other source.

APPENDIX

CHAPTER II OF THE INTRODUCTION TO KĀŞĞARĪ'S DĪWĀN LUĞĀTI'L-TURK*

THE LETTERS OF WHICH WORDS ARE BUILT UP

The letters by which the Turkish languages are rendered (*tadūru*) are eighteen basic (*aṣlīya*) letters in all, used in Turkish writing (*al-kitāba*). They are collected in the following mnemonic sentence (*al-maqāla*):

axūk laf samac nazaq baḏar ṣatīyā.†

Their isolated forms (*muqaṭṭa'ātuhā*) are written in the following fashion‡:—

aleph (*ā*), *beth* (*f*; *v* is intended), *gimel* (*x*), *vau* (*w*), *zain* (*z*), *cheth* (same outline as *gimel* with two superscribed dots; *q*),

* Facsimile, page 6; printed Arabic text, page 7; *Atalay*, I, 8.

† These seem to be intended to be Arabic words, "thy brother" and so on, but the vocalization is probably not that intended by Kāşğarl, who left them unvocalized.

‡ I have substituted the old Semitic names for the Uyğur letters, which are easily recognizable.

iod (*y*), *caph* (post-palatal *k*), *lamed* (*d̲*), *mem* (*m*), *nun* (same outline as *aleph* with one superscribed dot; *n*), *samech* (*s*), *pe* (*b*), *tzaddi* (*c*; *ç* intended), *resh* (*r*), *schin* (same outline as *samech* with two subscribed dots; *ṣ*), *tau* (*t*), hooked *resh* (*l*).

These letters correspond to the Arabic letters *alif*, *bā tā*, *ṯā*.

There are seven other, subsidiary (*far'īya*), letters of which there is no mention (*ḏikr*) in the script (*al-xaṭṭ*), and words are not distinguished (*yanfakk*) by them.

These are the hard *bā* (*al-bā'u'l-ṣulba*; **p**); the Arabic *cīm* (**c**), which is rare in this language; *zāy* pronounced between the points of articulation (*maxracay*) of *zāy* and *ṣīn* (i.e. **j**); the Arabic *fā* (**f**); *ǧayn* (**ğ**); *kāf* pronounced between the points of articulation of *qāf* and *kāf* (i.e. **g**); and the nasal *kāf* (*kāfu'l-ǧunna*) pronounced between *ǧayn* and *qāf* and *nūn* and *qāf*; this last letter is difficult for anyone who is not a Turk to pronounce.

These subsidiary letters are written in the shapes (*ṣīǧa*) of the basic letters but dotted (*tunqaṭ*), and are recognized by them (i.e. the dots).

The following (letters) are completely lacking in the Turkish languages:—*ṯā*; also of the emphatic letters (*ḥurūfu'l-iṭbāq*) *ṭā*, *ẓā*, *ṣād*, *ḍād*; also of the guttural letters (*ḥurūfu'l-ḥalq*) *ḥā*, *hā*, *'ayn*; it is true that the owl is called **ühi:**, but the correct word (*ṣamīmu'l-luǧa*) is **ügi:** with a *kāf*, a Kıpçak word; similarly a "strike-a-light" (*al-zand*) is called **caha:**, a voiced (*rakīka*) Gancak* word; and "ophthalmia" is called **awah** (so vocalized), and this too is not correct.

Apart from this *hā* is used "in pause" (*tadxul li'l-waqf*); for example one says **ta:h ta:h** in calling a falcon and **kurrıh kurrıh** in calling a foal, but that they are words containing **h** is not the case (*fa-ammā an yakūn kalāma*(*n*) *muhawwa'ata*(*n*), *fa-lā*).

* This word, with initial and final *kāf*, occurs fairly often in Kāşğarī as the name of a non-Turkish language. As it is used in Arabic, and not in Turkish phrases, the usual transcription *Kençek* is obviously a solecism. I take the name to be Iranian; it occurs as the name of a place in Azerbaijan, Ganja, later Elisavetpol, and now Kirovabad, but that is obviously a different place.

And *hā* is found in words of (the language of) Xotan, because it is an offshoot (*min natīca*) of India; and in words of (the language of) Gancak also, because it is not Turkish.

If it is necessary to write *ṭā* it is written in the shape of Turkish *ḏāl* and dotted above.

Similarly *ḍad* is written as *ḏāl* and dotted.

Ṣād is written as *sīn* and dotted.

But *ḥā* and *'ayn* and *hā* are written in the shape of *xā* and marked with a mark (*tu'lam 'alayhā bi-'alāma*). These letters are peculiar to Arabic, and if they are written in Turkish script (*hicā*), they resemble these letters in shape (*fī'l-qālib*) and are differentiated (*yufarraq*) by dots or marks.

The whole alphabet when written in combination (*cumi'at*) is as follows:—(in Uyğur script) *avaxa:*, *wasaka:* (*samech* is written in place of *ẓain*), *yekde:*, *mana:ẓ* (*ẓain* is written in place of *samech*), *baçara:*, *ṣatal a:*.

The principle (*al-aṣl*) in writing this script is that short a/e (*fatḥa*) requires (*taqtarif*) an *alif* (i.e. *aleph*) in the text without its having a function in the pronunciation (*aṣl fī'l-lafẓ*); and every short rounded vowel (*ḍamma*) procures (*tactariḥ*) a *wāw* (i.e. *vau*) in the script (*al-saṭr*) without its having a function in what is said (*fī'l-nuṭq*); and every short é/ı/i (*kasra*) gains (*taktasib*) a *yā* (i.e. *jod*) without its having a function. There is a similar practice (*naẓīruhā*) in Arabic in the words *al-ab* (" father ") and *al-ax* (" brother ") in the forms with possessive suffixes; one says (*sic*, but " writes " is meant) *hāḏā abūk*, *ra'aytu abāk*, and *marartu bi-abīk*; you write the word with these vowel letters in place of vowel signs (*al-ḥaraka*).

The rescripts and dispatches (*kutub wa marāsīl*) of the Xakans and Sultans have been in this script from ancient times to the present day from Kāşğar to Upper China (*al-Ṣīnu'l-'ulyā*) everywhere in all the lands of the Turks.

Turkish words contain back, front and modified vowels (*al-iṣbā' wa'l-imāla wa'l-iṣmām*) corresponding to the three vowel signs; and in the pronunciation (of consonants) unvoiced, voiced and nasal quality (*al-ṣalāba wa'l-rikka*

wa'l-ğunna), the guttural nasal letter (*al-ḥarfu'l-xayšūmī*), two consecutive consonants without an intervening vowel (*al-cam' bayna'l-sākinayn*), the collocation of *qāf* with a following *cīm*; and alternations between *bā* and *mīm* and between *nūn* and *lām* and others. Reference will later be made to these points.

CHAPTER VI

THE STRUCTURE OF THE WORD IN PRE-EIGHTH CENTURY TURKISH

An essential preliminary to the reconstruction of the phonetic structure of pre-eighth century Turkish is an analysis of the structure of the Turkish word. There is no reason to suppose that any fundamental change in this structure had taken place at any rate for some centuries before the eighth century, and we can confidently use the evidence of the eighth to eleventh centuries to establish this structure.

It was pointed out in Chapter II that there were (and still are) in Turkish two kinds of words, words carrying a suffix or suffixes of one kind or another, and words from which no known suffix can be detached and which can therefore best be described as " basic words." It was, to use an analogy, these basic words that were the bricks and the suffixes that were the mortar out of which the Turkish language was constructed. Later this native material was in all Turkish languages, but in some more than in others, supplemented by loan words, and in a few languages even loan suffixes, but this supplementary material is not taken into account in this chapter.

Basic words fall into two classes, verbs and words which are not verbs. The latter class includes words belonging to several grammatical categories, nouns, pronouns, numerals and so on, but as the overwhelming majority of them are nouns, and as there are no structural differences between words belonging to these various categories, except that double consonants occur only in a few numerals, and one or two onomatopoeics (which can hardly be regarded as words in the grammatical sense), all these words can conveniently, if a little inaccurately, be called " nouns " in antithesis to verbs.

In Turkish, unlike some other languages, there was one conjugational form, the second person singular of the imperative, which carried no suffix, and so can be used as the basic form of the verb. To distinguish between nouns and verbs I have adopted the convenient convention of attaching a hyphen to the latter; thus **at** " horse," **at-** " to throw."

A basic word consists of a single vowel sound (V) or a number of consecutive sounds, either vowels (V) or consonants (C). Long vowels (Vv) were relatively common in early Turkish, but double consonants (Cc), as already stated, occurred only in the numerals **ékki:, yétti:, sekkiz, tokkuz, ottuz** and **éllig** and in a few onomatopoeics.

Two consecutive vowels (VV) never occurred in early Turkish. Two consecutive consonants (CC) never occurred at the beginning of a word, and in basic words only certain consonants occur as the first or second member of such a pair medially and finally and only in combination with certain other consonants. The number of pairs of consonants which might occur at the end of any word, basic or suffixed, was strictly limited. The combination VvCC probably never occurred, although Kāşğarī (*I*, 341, 9) does mention **tö:rt** as a rare form of **tört** " four." A few apparently basic words contain three consecutive consonants (CCC), but it is very possible that the words are not in fact basic and that the third consonant is the first sound in an obsolete suffix.

One of the phonetic differences between basic and suffixed words was that in the latter certain consonants (the last consonant of the word and the first of the suffix) occurred in juxtaposition with one another which never occurred in juxtaposition within a basic word. The reason was that both basic words and suffixes were, phonetically speaking, exceptionally stable, and that assimilation (what is called in Sanskrit grammar *sandhi*) between the final consonant of a basic word and the initial consonant of a suffix seldom occurred. It was not however unusual to

insert a euphonic vowel between two such consonants if they were more than usually incongruous.

It has been suggested that in the most primitive form of Turkish, that is when words first crystallized out of less organized sounds, the first words which were invented or accepted were all monosyllables, and that the rest of the vocabulary was created by elongating this primaeval stock of monosyllables. If that were so, then the only truly basic words in the language would be the monosyllables and the second syllable of every dissyllable would have to be explained as a suffix. This theory cannot be maintained. It is true that there are a good many monosyllables in early Turkish, but taken together they would be quite insufficient to provide a viable vocabulary; for example the only monosyllabic numerals are those meaning one, three, four, five, ten, forty, a hundred and a thousand. On the other hand, there are many dissyllables and a few trisyllables from which it is quite impossible to detach any known or hypothetical suffix so as to isolate the hypothetical basic monosyllable.

There seems in basic words to have been a sort of natural affinity between short vowels (V) and subsequent unvoiced consonants, and long vowels (Vv) and subsequent voiced consonants. If an initial or medial V, at any rate in a monosyllable, was followed by a plosive, fricative, affricate or dental sibilant that consonant seems always to have been unvoiced. If an initial or medial Vv, at any rate in a monosyllable, was followed by such a consonant it seems to have been almost invariably voiced. The few apparent exceptions require further investigation. On the other hand, both V and Vv seem to precede nasals, liquids, ş and y.

Further study is also required on the effect which the attachment of a suffix had on such a preceding long vowel. The only reliable evidence on this subject is provided by Kāşğarī, but, as said in Chapter V (1) (vi), it is difficult to interpret. There are, however, some indications that the

attachment of a suffix, at any rate in certain circumstances, shortened a preceding long vowel. The reason was presumably a shift of stress.

The following forms of basic monosyllabic nouns and verbs occur:—

Vv (naturally very few)	CVv
VC	CVC
VvC	CVvC
VCC	CVCC*

There are no known cases of V or CV.

As there are some differences of structure between basic dissyllabic nouns and verbs, it will be better to treat them separately.

The forms of basic nouns were:—

—	CVCV (only **Küli** and **Şiri**?)
VCVv	CVCVv
VvCVv (only **ü:gi:/ü:hi:** "owl"?)	—
VCcVv (only **ékki:** and onomatopoeics?)	CVCcVv (only **yétti:** and onomatopoeics?)
—	CVCCV (only **Türkü**?)
VCCVv	CVCCVv
VCVC	CVCVC
VCcVC (only **ottuz, éllig** and onomatopoeics?)	CVCcVC (only **sekkiz, tokkuz** and onomatopoeics?)
VCVvC	CVCVvC
—	CVvCVC
—	CVvCVvC (only **ka:ğu:n**?)
VCCVC	CVCCVC
VCCVvC	CVCCVvC

Of these by far the commonest were V and CVCVC; VCVv, VCCVC and CVCCVv were relatively common; the rest were rare or very rare. There were also one or two words of such forms as VCCCVvC and

* Perhaps also CVv CC, but this is improbable.

CVCCCVvC, but these were perhaps not really basic words. In the aggregate there were, as might be expected, a good many more basic dissyllabic nouns than monosyllabic.

There were also a few trisyllabic nouns like **yégirmi:** " twenty " from which it does not seem possible to detach a suffix, but so few that it is unnecessary to enumerate their forms. There do not appear to be any trisyllabic words containing CCC which can be confidently classified as basic.

The forms of basic dissyllabic verbs were substantially fewer. The commoner forms were:—

VCVv-	CVCVv-
VCCVv-	CVCCVv-

The following were relatively rare:—

—	CVvCVv-
VCVC-	CVCVC-
VCCVC-	—

It looks as if there were more dissyllabic basic verbs than monosyllabic, but not many more.

There were no basic trisyllabic verbs.

CHAPTER VII

THE SUFFIXES IN PRE-EIGHTH CENTURY TURKISH

In the last chapter we considered the structure of the basic word; we now have to consider the other class of components out of which the Turkish language was constructed, the suffixes which were attached to basic words and must be peeled off one by one to reach them. The method of making new words by elongating existing ones has always been, and still is, one of the most characteristic features of the Turkish language. It is, however, clear that just as there has been a slow secular change in the Turkish vocabulary, so that words which were common in the eighth century have now been forgotten, and words which are now common did not then exist, so also there has been a slow secular change in the suffixes used to make new words. Some words in the eighth century vocabulary carry suffixes which were no longer "productive," that is could no longer be attached to other words to form new words. Indeed some such suffixes had even then been so long obsolete that it is difficult to guess what their original functions were, although we can be certain that they were suffixes. Some suffixes which were still productive in the eleventh century, for example the suffix **-sa:-/-se:-** to form the desiderative verb, are no longer productive. Kāşğarī (*I*, 279 ff.) devotes some space to explaining this desiderative suffix, but even if we knew nothing of any form of Turkish more than a hundred years old, it would still be possible, by comparing Osmanli **su** "water" and **susa-** "to be thirsty," and one or two other similar pairs of words to prove that there must at one time have been a desiderative suffix of this form. Other suffixes, for example **-tur-/-tür-** and **-VI-**, the suffixes used to form causative and passive verbs respectively, were productive in the eighth century and are still productive. And finally there are some suffixes

which are productive to-day but seem to have come into use only comparatively recently.

One proof of the fact that some of the suffixes which can be identified in words of the eighth to eleventh century vocabulary had long been obsolete, and that these words go back to a much earlier state of the language, is that when the suffix in question is peeled off what is left is a presumably basic word, which was no longer in current use and sometimes was even in a form which was no longer current. For example the earlier existence of a nominal suffix **-ğa:/-ge:** is proved by its presence in **bilge:** " wise; Chancellor " (the title of an office) and **öge:** " Adviser " (the title of another office), derived respectively from **bil-** " to know " and **ö:-** " to remember." But when the same suffix is peeled off **avıçğa:** " old man," what is left is presumably a basic verb **avıç-** of which there is no other trace, and the very form is without parallel, since no verb of the form VCVC- seems to be known in which the second C is **-ç-**. It is however useful to compile a list of unproductive suffixes, as well as of those known at a time when they were still productive, since this will give us a deeper insight into the history of the language.

Suffixes are of various kinds, and the English language, which is structurally very different from Turkish, is not very well provided with technical terms for describing them. They can however be divided into two, or perhaps three, main classes.

Working forward from the basic words which can, or at one time could, be used in unsuffixed form as components in a sentence—the nouns proper as the subject of the sentence and in other appropriate ways, the verbs as imperatives and the other parts of speech with their appropriate functions—the first main class of suffixes comprises those which can be attached to a basic word in order to form an elongated word—noun, verb or other part of speech—which can be used as a component in a sentence in exactly the same way as a basic word of the same class. These can be called " primary suffixes " to distinguish them

from the rest. The term is not wholly satisfactory since it is often possible to add an additional primary suffix to a word which already carries one or more primary suffixes and still form a word which can be used as a component in a sentence in exactly the same way as a basic word.

The second main class of suffixes comprises those suffixes which are attached to basic words and words carrying primary suffixes in order to modify their meanings and functions in the sentence. These include possessive and declensional suffixes attached to nouns, conjugational suffixes attached to verbs and the like. It should be added that some suffixes have, in different circumstances, both the quality of a primary and the quality of a secondary suffix, for example some conjugational suffixes are also used to form verbal nouns. These secondary suffixes and their functions are fully discussed in the various Turkish grammars, for example, in the case of early Turkish, in von Gabain's *ATG*, and I need not refer to them further.

It is a matter of terminology whether the third class of syllables (or words) should be called suffixes or postpositions. They resemble suffixes in having no separate existence and being necessarily attached to some other word; they differ from them in that they are usually written as separate words, often modify the meaning not only of the word to which they are attached, but also of phrases or even whole sentences, and in some cases, for example when attached to pronouns, require the interposition of a declensional suffix. On the whole the differences outweigh the similarities, and postposition is the more appropriate description. They too are fully discussed in the Turkish grammars and I need not refer to them further.

One word of warning is necessary. Some sounds or combinations of sounds which are sometimes used as primary suffixes are at other times an integral part of a basic word and cannot be detached from it. There can be no fixed rule for distinguishing between the two cases; it is simply a matter of judgment, and in some cases it may

not be possible to reach general agreement whether a particular word is a basic or an elongated word. One such sound is final -z. Everyone would agree that it is a suffix in **ékkiz** "a twin" derived from **ékki:**. It is hard to believe that anyone would seriously contend that it is a suffix in **tokkuz** "nine" and **ottuz** "thirty." Between these two extremes is a whole list of words ending in -z which are names of parts of the body:—**ağız** "mouth," **omuz** (**omoz?**) "shoulder," **boğuz** (**boğoz**) "throat," **beŋiz** "complexion," **ti:z** "knee," **kögöz** "breast," **kö:z** "eye," **yamız** (or **yemiz?**) "the inner part of the thigh," and perhaps others. These look like basic words, since if the -z is peeled off no recognizable basic words are left, but the coincidence that so many words meaning parts of the body should end in -z is very odd; it may be that at some very early period when the basic vocabulary was being evolved it was the fashion to form words with this kind of meaning in this way. But the theory that -z is an old suffix for the dual which has been put forward has little to commend it. It seems to be based on nothing more substantial than that this final occurs in the names of some parts of the body which occur in pairs, like **kö:z** "eye," and that **ékkiz** means "a twin," but no one has two mouths or throats, and in **ékkiz** the duality is inherent not in the suffix but in the basic word **ékki:**; if -z really was a suffix for the dual then **ékkiz** ought to mean "four."

The following is a rather tentative list of the primary suffixes which seem, from an examination of the vocabularies of Türkü, Uyğur and Xākānī, to have existed in pre-eighth century Turkish. It is tentative not in the sense that I doubt whether some of the suffixes included in it ever existed, but in the sense that I do not feel confident that I have included in it all the suffixes which at one time existed but had become unproductive by the eighth century. I have tried to distinguish in it between those suffixes which were already unproductive by the eighth century, those which have since become unproductive and those which are still productive. The suffixes have been divided into

classes and sub-classes with a separate list for each arranged in alphabetical order (as set out in Chapter III) working backwards from the final sound. My purpose in compiling these lists is to provide an instrument for breaking down early Turkish words into their component parts, isolating the basic words and so deepening our knowledge of the Turkish language. I have briefly described the function of each suffix, where this can be discerned, and in some cases added references to other works in which they are explained in greater detail.

This is not of course the only list of its kind; there are, for example, such lists in von Gabain's *ATG* (here cited as *ATG* followed by a paragraph number) and C. Brockelmann, *Osttürkische Grammatik*, Leiden, 1954 (cited as *B* followed by a paragraph number), but these lists were compiled for different purposes and on different principles. In particular I have only occasionally listed separately compound suffixes, that is a primary suffix of one sub-class attached to another primary suffix of the same or a different sub-class.

Primary suffixes can be divided into three classes: (1) "nominal suffixes" used to make elongated nouns, adjectives and adverbs; (2) suffixes attached to numerals and the like; (3) "verbal suffixes" used to make elongated verbs; and (1) and (3) can be further subdivided into (a) "denominal suffixes," that is those attached to basic or elongated "nouns" (in the broader sense of the word used in Chapter VI) and (b) "deverbal suffixes," that is those attached to basic or elongated verbs.

(1) NOMINAL SUFFIXES

(a) DENOMINAL

-ça:/-çe: Diminutive (*ATG* 45) and Equative (*ATG* 349); perhaps to be distinguished from one another, since in modern languages the first is stressed and the second unstressed; still productive with additional connotations; in later stages of the language a homophonous

Persian suffix entered the language which is sometimes confused with this suffix.

-la:/-le: apparently adjectival, e.g. **körk** " beauty " > **körkle:** " beautiful "; rare and unproductive.

-ğıña:/-giñe:/-kıña:/-kiñe: forms diminutive adjectives and adverbs (*ATG* 351); probably productive till about the eleventh century.

-ra:/-re: " in "; perhaps originally a declensional suffix (*ATG* 187, 394, 429); e.g. **iç** " the inside " > **içre:** " within "; rare and unproductive; cf. **-ru:/-rü:**.

-çı:/-çi: Noun of Agent; common and still productive.

-tı:/-ti: forms adverbs; e.g. **edgü:** " good " > **edgüti:** " well " (*ATG* 382); rare and unproductive; also a conjugational suffix.

-tırtı:/-tirti:/-turtı:/-türti: forms adverbs of place, e.g. **öŋ** " front, east " > **öŋtürti:** " in front, in the east " (*ATG* 382); rare and unproductive.

-kı:/-ki: adjectival, " situated in," e.g. **içre:** " within " > **içre:ki:** " situated within, domestic " (*ATG* 49, 74); still productive, but now normally attached to genitive or locative.

-du:/-dü: function uncertain; e.g. **ka:r** " snow " > **kardu:** " hailstone " (*B* 43); rare and unproductive.

-ğu:/-gü: apparently forms abstract nouns from adjectives; e.g. **énç** " peaceable " > **énçgü:** " peace "; rare and unproductive; also a deverbal suffix.

(**-ŋu:/-ŋü:** See (c) Deverbal.)

-ru:/-rü: " into "; perhaps originally a declensional suffix; e.g. **e:v** " house " > **evrü:** " into the house " (*ATG* 429); rare and unproductive; not to be confused with gerund suffix **-u:/-ü:** attached to verbs ending in **-r-**.

-ğaru:/-gerü: synonymous with **-ru:/-rü:**; e.g. **iç** " the inside " > **içgerü:** " inwards "; commonly explained as a compound suffix, **-ru:/-rü:**, attached to the dative, but this can hardly be correct since this suffix already existed at a time when the dative suffix was still **-ka:/-ke:** and had not yet evolved into **-ğa:/-ge:**.

-layu:/-leyü: properly the gerund of a denominal verb, but the only form in which such verbs are recorded; usually, perhaps always, attached to the name of an animal; e.g. **arsla:n** " lion " > **arslanlayu:** " like a lion "; rare, but perhaps productive till the eleventh century.

-Vç Diminutive and affectionate; e.g. **ata:** " father " > **ataç** " dear father " (*ATG* 44); rare and unproductive.

-dıç/-diç function uncertain; e.g. **sa:ğ** " sincere " *etc.*; > **sağdıç** " close friend "; very rare and unproductive.

-ğVç/-gVç function uncertain; e.g. **o:d** " fire " > **odğuç** " firebrand," **kuş** " bird " > **kuşğaç** " sparrow "; rare and unproductive; also a deverbal suffix.

(**-la:ç/-le:ç** See (c) Deverbal.)

(**-ğut/-güt** Probably Deverbal; see (c).)

-çığ/-çig apparently a secondary form of **-sığ/-sig,** possibly a crasis of ***-çsığ/-çsig**; see *ATG* 80; also a deverbal suffix.

-ta:ğ/-te:g (or **-da:ğ/-de:g**?) properly the postposition **te:g,** but fused with a basic word in such words as **anta:ğ** (or **anda:ğ**?) " like that "; rare and unproductive in fused form.

-lığ/-lig forms possessive adjectives; common and still productive, usually in altered forms (**-lı/-li** *etc.*).

-sığ/-sig properly deverbal **-ğ/-g** attached to a denominal simulative verb which is not used as such; e.g. **er** " man, male " > **ersig** " manly "; rare and unproductive.

-Vk usually diminutive and affectionate; e.g. **ö:g** " mother " > **ögük** " little mother " (*ATG* 57); rare and unproductive.

-çak/-çek occurs in a number of names of concrete objects; some have no obvious etymology, but **bağır** " liver, loins " > **bağırçak** " donkey's pack saddle "; unproductive and probably very old; also a deverbal suffix.

-çuk/-çük described by Ḳāşğarī as a diminutive; still productive, but in the eleventh century a rare new suffix.

-dak/-dek function uncertain; e.g. **bağır** " liver, loins " > **bağırdak** " body garment "; rare and unproductive.

-duk/-dük/-tuk/-tük function uncertain; e.g. **kol** " arm " > **koltuk** " arm-pit "; very rare and unproductive; perhaps abbreviation of **-duruk** *etc.*; a common conjugational suffix.

-ğak/-gek a fairly common deverbal suffix, but apparently sometimes denominal; possibly a very old suffix, but some examples in *ATG* 59 look more like basic words.

-lık/-lik forms abstract, and less often concrete, nouns; common and still productive.

-(V)muk/-(V)mük function uncertain, possibly pejorative; e.g. **kara:** " black " > **kara:muk** " a kind of weed," and **sol** " left " > **sola:muk** " left-handed "; rare and unproductive.

-ñak/-ñek occurs in a few words, most with no obvious etymology, but **baka:** " frog " > **baka:ñak** " the frog in a horse's hoof "; rare and unproductive.

-rak/-rek comparative form of adjective; productive until the fifteenth century and perhaps later, but now unproductive.

-duruk/-dürük forms names of implements; e.g. **boyun** " neck " > **boyunduruk** " a yoke "; probably productive till the eleventh century.

-sak/-sek properly deverbal **-k** attached to desiderative verb, but in some cases the verb itself is not recorded; e.g. **er** " man " > **ersek** " nymphomaniac "; probably productive till the eleventh century; borrowed as a suffix by Mongolian.

-suk/-sük function uncertain; e.g. **bağır** " liver, loins " > **bağırsuk** " entrails "; rare and unproductive.

-Vl possibly forms adjectives; e.g. **yaş** " vegetation " > **yaşıl** " green "; rare and unproductive; also a deverbal suffix.

-çıl/-çil forms adjectives, usually frequentative or connoting addiction; e.g. **ig** " disease " > **igçil** " sickly,"

and **yağmur** " rain " > **yağmurçıl** " rainy "; perhaps still productive in the eleventh century and even later, but now unproductive.

-ğıl/-gil apparently associated with colour; e.g. **baş** " head " > **başğıl** " grey-haired," and **yal** " mane " > **yalğıl** " white maned "; rare and unproductive; also a numeral suffix.

-sıl/-sil function uncertain; both **a:r** and **arsıl** mean " auburn "; rare and unproductive; possibly associated with **-çıl/-çıl.**

-Vm function uncertain; e.g. **eder** " saddle " > **edrim** " saddle-pad "; rare and unproductive, but still common as deverbal and possessive suffix.

-dam/-dem " resembling "; usually adjectival, e.g. **teŋri:** " God " > **teŋridem** " god-like "; but also forms nouns, e.g. **er** " man " > **erdem** " manliness "; rare and unproductive.

-Vn obsolete plural suffix; so far noted only in **er** " man " >**eren** " men "; **oğul** " son " > **oğla:n** " sons "; **bo:d** " clan " > **bodun** " tribe "; **ö:z** (in two meanings) > **özen,** and possibly **o:y** " hole " > **oyun** " game "; the plural connotation was already forgotten in the eleventh century.

-çın/-çin probably a secondary form of **-çıl/-çil**; e.g. **balık** " fish " > **balıkçın** " fish-eating bird "; rare and unproductive.

-dun/-dün (**-tun/-tün**) " situated in "; e.g. **öŋ** " front " > **öŋdün** " in front " (*ATG* 183); possibly an old declensional suffix but distinct from the ablative suffix **-dın** *etc.*

-ğa:n/-ge:n function uncertain; e.g. **arpa:** " barley " > **arpağa:n** " wild barley," and **yétti:** " seven " > **yétige:n** " the Great Bear "; also a common deverbal suffix.

-ğun/-gün apparently a collective: e.g. **alku: (alko:)** " all " > **alkuğun** " all together "; very rare and unproductive.

-kan/-ken forms adjectives connoting addiction (?);

e.g. **teŋri:** "God" > **teŋriken** "devout"; very rare and unproductive.

-la:n/-le:n possibly a very old unproductive suffix; occurs in a number of names of animals with no obvious etymology, e.g. **arsla:n, bursla:n, kapla:n, kula:n, yıla:n** (***d̲ıla:n**).

-man/-men function uncertain; e.g. **kö:z** "embers" > **közmen** "bread baked in the embers"; as such rare and unproductive, but in some later languages apparently a secondary form of Persian suffix *-mand* "resembling"; this later suffix may be the one in **Türkü** > **Türkmen.**

-şı:n/-şi:n probably a secondary form of **-çın/-çin** (**-çıl/-çil**); e.g. **kök** "blue" > **kökşi:n** "bluish"; rare and unproductive.

-Vŋ function uncertain; e.g. **kö:l** (**gö:l**) "lake" > **kölüŋ** "puddle"; also seems to occur in some other words ending in **-ŋ** with no obvious etymology, if so a very old suffix, unproductive; also a possessive suffix.

-dVŋ/-tVŋ function uncertain; occurs in **o:d** "fire" > **otuŋ** (for **odtuŋ**) "firewood"; **u:l** "foundation" > **uldaŋ** "sole"; **a:y** "moon" > **aydıŋ** "moonlight"; and perhaps **izdeŋ** "a kind of fish-net."

-Vş function uncertain; e.g. **ba:ğ** and **bağış** both mean "tie, fastening"; rare and unproductive but a common deverbal suffix.

-da:ş/-de:ş "sharing, partners in (something)"; e.g. **karın** "womb" > **karındaş** "brother"; has been explained rather plausibly as crasis of locative suffix **-da:/de:** and **éş** "companion"; still productive.

-ğay/-gey function uncertain, possibly adjectival; e.g. **küç** "strength" > **küçgey** "violent"; rare and unproductive; also a conjugational suffix.

-Vz function uncertain; e.g. **ékki:** "two" > **ékkiz** "twin"; see the remarks above; rare and unproductive.

-duz/-düz function uncertain; e.g. **kün** "sun, day" > **kündüz** "daytime"; occurs in other words like

kunduz " beaver " which have no obvious etymology; perhaps a very old suffix, unproductive.

-sız/-siz " not possessing," opposite to **lığ/lig**; still productive.

(b) SUFFIXES ATTACHED ONLY TO NUMERALS AND THE LIKE*

-nti: occurs only in **ékki:** " two " > **ékkinti:** " second "; perhaps the original ordinal suffix.

-ağu:/-egü: collective suffix; apparently attached to one or two cognate words, e.g. **ad̲ın** " other " > **ad̲ınağu:** " others."

-Vnç ordinal suffix, perhaps a crasis of **-nti:**.

-ğıl/-gil occurs in **üç** " three " > **üçgil** " triangular " and **tört** (**dört**) " four " > **törtgül** " rectangular "; also a denominal suffix.

-Vr and **-şa:r/-şe:r** distributive suffixes.

-mış/-miş occurs only in **altmış** " sixty " and **yétmiş** " seventy "; apparently a very old suffix for " ten."

(c) DEVERBAL

-a:/-e:/-ı:/-i:/-u:/-ü: these are all gerundial suffixes and it is not easy to distinguish between normal gerunds and words which have acquired other connotations; an intermediate class is constituted by words like **u:d-** " to follow " > **ud̲u:** " thereafter " and **bas-** " to press " > **basa:** " then " which function as adverbs; but there are a few clear cases of nouns formed with this suffix, e.g. **ad̲ır-** " to divide " > **ad̲rı:** " pitch-fork "; rare and unproductive.

-ğa:/-ge: forms nouns of action or state, e.g. **bil-** > **bilge:** and **ö:-** > **öge:** mentioned above, but a good many words with this suffix have no obvious etymology, e.g. **avıçğa:** " old man "; a very old and rather rare suffix, unproductive.

-ma:/-me: a suffix only beginning to come into use in the

* On this subject see *The Turkish numerals*.

eighth century and still rare in the eleventh, at that time usually with a passive connotation, e.g. **kes-** " to cut " > **kesme:** " forelock," but now the ordinary noun of action in languages like Osmanli/Republican Turkish; still productive.

-ğma:/-gme: properly a participial suffix, but sometimes used to form adjectives and even nouns; productive probably till the eleventh century, but not much later; cf. **-ğlı:/-gli:**.

-çı:/-çı: properly a denominal suffix, but apparently in the eleventh century sometimes attached to verbs; e.g. **okı:-** " to call " > **okı:çı:** " pleader "; rare and unproductive.

-daçı:/-deçi:/-taçı:/-teçi: } these suffixes (the latter a com-
-ğu:çı:/-gü:çi: } bination of **-ğu:/-gü:** and the denominal suffix **-çı:/-çi:**) both form nouns of agent or state, also used as adjectives and practically indistinguishable from participles; according to Kāşğarī (*II*, 47 ff.) the first was the western (Oğuz, Kıpçak, *etc.*), and the second the eastern (Uyğur, Xākānī, *etc.*) form; productive till the eleventh century but not much later.

-dı:/-di: function uncertain in words like **ö:g-** " to praise " > **ögdi:** " praise "; commoner when attached to reflexive verbs, when it functions as an adjective or noun usually with a passive connotation, e.g. **üdrün-** " to be chosen " > **üdründi:** " chosen, select "; probably still productive attached to reflexive verbs; also a conjugational suffix.

-ğlı:/-gli: practically synonymous with **-ğma:/-gme:** and, like it, properly a participial suffix.

-mı:/-mi: function uncertain; noted only in ***tegir-** " to revolve, be round " (?) > **tegirmi:** " circular, round."

-çu:/-çü: Kāşğarī lists nearly a dozen words with this suffix derived from reflexive verbs with much the same connotation as words in **-dı:/-di:** similarly derived, e.g. **atin-** " to be thrown " > **atınçu:** " thrown away, discarded "; rare and unproductive.

-ğu:/-gü: forms noun/adjectives with varying functions (*ATG* 115, 141); common and productive at any rate until the eleventh century; also a denominal suffix.

-ŋu:/-ŋü: in some words like **tayan-** "to rely on" > **tayaŋu:** "minister," clearly a crasis of **-n-** and **-ğu:/-gü:**, but there are one or two words like **esri:** > **esriŋü:** "parti-coloured, dappled" which cannot be so explained, possibly denominal.

-p a common gerundial suffix; *ATG* 126 suggests that three monosyllables ending in **-p**, e.g. **kop** "all" are gerunds used as adjectives, but the etymologies are dubious and there are no similar dissyllables.

-ç forms nouns, usually of state or condition, from reflexive verbs; e.g. **ögrün-** "to be joyful" > **ögrünç** "joy"; probably a crasis of **-ış/-iş** which seems not to be attached to reflexive verbs; this suffix is sometimes listed as **-nç** but this seems to be an error; productive at any rate till the eleventh century but now unproductive.

-ğVç/-gVç usually forms names of instruments; e.g. **kıs-** "to grip" > **kısğaç** "tongs"; probably productive till the eleventh century; also a denominal suffix.

-laç/-leç seems to be a suffix in one or two words with no obvious etymology like **ıkı:la:ç** "a spirited horse" and **sukarlaç (börk)** "a tall cap"; unproductive and very old; possibly denominal.

-ma:ç/-me:ç forms names of kinds of food; e.g. **tut-** "to hold" > **tutma:ç** "noodles"; perhaps productive till the eleventh century, but rare.

(**-nç**. See **-ç**.)

-Vt usually forms nouns of action; e.g. **adır-** "to distinguish" > **adırt** "distinction," and **bas-** "to press," *etc.* > **basut** "help"; rare and probably unproductive by the eleventh century.

-ğut occurs in three words, **alp** "tough" > **alpa:ğut** "a brave fighter," **ba:y** "rich" > **baya:ğut** "a rich merchant" and **ura:ğut** "woman" with no obvious etymology; these words are best explained as **-ğut**

attached to a denominal verb in **-a:-**, since **-a:-** seems to be an integral part of the suffix and not a euphonic vowel, and no dissyllabic suffixes (except **-ğıña:** *etc.*) seem to be known; a very old suffix, rare and unproductive.

-Vğ/-Vg the commonest suffix of deverbal nouns, usually of action but with other shades of meaning; the difference, if any, between this suffix and **-Vk** is obscure; possibly still productive (in secondary form -Vv) in some modern languages.

-çığ/-çig apparently adjectival; attached only to reflexive verbs, e.g. **yarsın-** "to be revolted" > **yarsınçığ** "revolting"; rare and unproductive; perhaps a crasis of **-ç** and **-sığ/-sig**; also a denominal suffix.

(**-nçığ/-nçig**. See **-çığ/-çig**.)

-kVğ/-kVg and **-ğVk/-gVk** these are perhaps different suffixes (see *ATG* 114 and *B* 52); at any rate sometimes connote habitual action, e.g. **iç-** "to drink" > **içgek** "vampire"; the phonetics of the whole group are obscure and some words carrying this suffix have no obvious etymology; unproductive and probably very old.

-Vk a common deverbal suffix, sometimes with passive connotation, e.g. ***anu:-** "to prepare" > **anuk** "ready"; the difference, if any, between this suffix and **-Vğ/-Vg** is obscure; still productive in the eleventh century but not much later.

-çak/-çek function uncertain; e.g. **urun-** "to put down" > **urunçak** "pledge, guarantee"; attached only to reflexive verbs; rare and unproductive; also a denominal suffix.

-dak/-dek function uncertain; e.g. **yalın-** "to strip oneself" > **yalındak** "naked"; rare and unproductive; also a denominal suffix.

-mVk probably the suffix of the infinitive, **-ma:k/-me:k** used to form nouns, but this is not certain since in such words the vowel is often short and subject to vocalic assimilation at any rate in the mediaeval

period; common (see *B* 84) but probably unproductive by the eleventh century.

-sık/-sik function uncertain, possibly a combination of simulative verbal suffix **-sı:-/-si:-** and **-k**; e.g. **bat-** "to set" > **kün batsıkı:** "sun-set"; (*ATG* 129, 151); rare and unproductive.

-yuk/-yük conjugational form (perfect) occasionally used as adjective or even noun (*ATG* 134, 152); e.g. **sı:-** "to break" > **sıyuk** "broken"; unproductive by the eleventh century.

-Vl function uncertain; e.g. ***ına:-** "to trust" > **ınal** "confidant"; rare and unproductive and probably very old; also a denominal suffix.

-ğıl/-gil function uncertain; e.g. **bıç-** "to cut" > **bıçğıl** "crack, cracked, split"; rare and unproductive; also denominal and conjugational (imperative) suffix; not to be confused with Mongolian suffix **-ğul/-gül** found later in Mongolian loan words.

-Vm common deverbal noun both abstract and concrete; usually connotes a single action or thing, e.g. **o:b-** "to swallow" > **obum** "a mouthful"; perhaps still productive.

-Vn function uncertain; e.g. **tüg-** "to knot" > **tügün** "a knot"; rather rare, but apparently productive in the eleventh century and perhaps later.

-ğa:n/-ge:n conjugational suffix (present participle) often used as adjective connoting frequentative action; some words carrying this suffix, like **tavışga:n** "hare," have no obvious etymology so it must be very old; productive until the eleventh century and perhaps later.

-ğın/-gin function uncertain; e.g. **kev-** "to masticate" > **kevgin** "indigestible"; rare and unproductive and probably very old.

-ğun/-gün forms both adjectives, e.g. **ol-** "to mature" > **olğun** "ripe," and nouns, e.g. **ört-** "to cover" > **örtgün** "stacked corn"; still productive in mediaeval period and probably not very old.

-maŋ/-men function uncertain; e.g. **ört-** "to cover" > **örtmen** "covering, rug"; rare and unproductive; also a denominal suffix.

-çaŋ/-çeŋ apparently frequentative; **unut-** "to forget" > **unutçaŋ** "forgetful"; rare and unproductive.

-r conjugational suffix (aorist participle) sometimes used to form nouns/adjectives; probably still productive.

-mVr function uncertain; e.g. **ya:ğ-** "to rain" > **yağmur** "rain"; rare and unproductive.

-Vş common suffix usually forming abstract nouns; e.g. **köse:-** "to wish" > **kösüş** "a wish"; still productive; in Osmanli/Republican Turkish often connotes the manner in which something is done.

-mış/-miş conjugational suffix (reported perfect participle) often used, according to *ATG* 122, as a predicative noun; now productive only as conjugational suffix.

-ma:z/-me:z conjugational suffix (negative aorist participle) sometimes used to form adjectives and nouns (*ATG* 121); probably still productive.

(2) VERBAL SUFFIXES

(a) DENOMINAL

-a:-/-e:- (**-ı:-/-i:-, -u:-/-ü:-**) usually forms transitive verbs; **-a:-/-e:-** is fairly common, the others rare; probably the oldest denominal suffix, unproductive by the eighth century.

-da:-/-de:- secondary form of **-la:-/-le:-** occasionally found attached to nouns ending in **-ğ**; e.g. **ba:ğ** "a tie" > **bağda:-** "to bind," **-l**, e.g. **a:l** "device" > **alda:-** "to trick, deceive," and **-n**, e.g. **ün** "voice" > **ünde:-** "to call"; now unproductive.

-ta:-/-te:- secondary form of **-la:-/-le:-** occasionally found attached to nouns ending in **-k**, e.g. **ok** "arrow" > **okta:-** "to shoot"; now unproductive.

-ğa:-/-ge:-/-ka:-/-ke:- over a dozen verbs in **-ğa:-/-ge:-** and rather fewer in **-ka:-/-ke:-** occur in Kāşgarī; hardly any have obvious etymologies; some may be

basic verbs or verbs formed by **-a:-/-e:-** attached to a deverbal noun in **-ğ/g** or **-k**, but **yarlıka:-** (***yarlığka:-**) "to command" seems to be derived from **yarlığ** "a command" and **irinçke:-** "to pity" from **irinç** "misery."

(**-Vrğa:-/-Vrge:-**. See **Vrğan-/-Vrgen-**.)

-la:-/-le:- the standard denominal suffix for forming transitive verbs from the eleventh century or earlier onwards; still productive.

-ra:-/-re:- there are some traces of denominal verbs with this suffix, see *ATG* 95, *B* 171; rare and unproductive.

-sıra:-/-sire:- "to lack, be without," cognate to, and etymologically connected with, nominal suffix **-sız/-siz** (*ATG* 99, *B* 173); rare and unproductive.

-sa:-/-se:- forms desiderative verbs; productive till the eleventh century but not much later; also a deverbal suffix.

-lı:-/-li:- occurs in **tum** "cold" > **tumlı:-** "to be cold" and perhaps one or two other verbs with no obvious etymology; rare and unproductive.

-sı:-/-si:- forms simulative verbs; very rare except in reflexive form attached to deverbal nouns in **-m** (see *ATG* 157, **-Vmsın-/-Vmsin-**) and not common even then; unproductive.

-ad-/-ed- forms intransitive and transitive verbs; e.g. **kut** "good fortune" > **kutad-** "to be, or make, fortunate"; rare and unproductive.

-Vk- forms intransitive or passive verbs; e.g. **iç** "interior" > **içik-** "to come in, submit," and **ağu:** "poison" > **ağuk-** "to be poisoned"; rare and unproductive; also a deverbal suffix.

-Vl- forms intransitive verbs; e.g. **tusu:** "utility" > **tusul-** (as well as **tusuk-**) "to be useful," and one or two other verbs; rare and unproductive; also a deverbal suffix.

-Vn- is a possible denominal suffix, but more likely to be the reflexive form of a denominal verb in **-a:-/-e:-**, *etc.*, see *ATG* 90.

-Vrğan-/-Vrgen- reflexive form of -Vrğa:-/-Vrge:-, but for practical purposes the only form in which this suffix occurs; forms intransitive verbs; e.g. **alp** "tough" > **alpırğan-** "to make strenuous efforts"; *ATG* 96 transcribes -Vrka:-/-Vrke:-, but this seems to be an error; rare and unproductive.

-Vr- forms intransitive verbs; e.g. **ürüŋ** "white" > **ürüŋer-** "to be white"; perhaps etymologically connected with -ad̲-/-ed̲-; rare and unproductive; also a deverbal suffix.

-ğar-/-ger- (-kar-/-ker-) apparently forms both transitive and intransitive verbs (*ATG* 87, *B* 163); e.g. **ot** "pasture" > **otğar-** "to pasture (animals)," and **köŋül** "mind" > **köŋülger-** "to think"; rare, perhaps productive till mediaeval period; *B* 164 also quotes forms with -ğır-/-gir-/-ğur-/-gür- but these seem to be deverbal forms.

?-Vş- *ATG* 101 quotes this as a denominal suffix but it seems rather to be deverbal.

(b) DEVERBAL

-ma:-/-me:- negative; common at all periods and still productive.

-sa:-/-se:- desiderative; productive until the eleventh century but not much later; also a denominal suffix.

-sı:-/-si:- simulative; e.g. **em-** "to suck" > **emsi:-** "to seem to suck"; rare but commoner in the reflexive form; probably no longer productive in the eleventh century; also a denominal suffix.

-d̲- existence and function uncertain, possibly "emphatic," deduced from **to:-** "to fill" > **tod̲-** "to be full" and perhaps ***ko:-** > **kod̲-**, see *ATG* 153; also a denominal suffix.

-t- causative suffix of verbs ending in vowels and sometimes -r-; common at all times and still productive.

-Vk- when attached to intransitive verbs "emphatic"; e.g. **a:c-** "to be hungry" > **acık-** "to be famished";

when attached to transitive verbs forms intransitive, e.g. **böl-** " to separate " > **bölük-** " to be separate "; rare and unproductive; also a denominal suffix.

-sık-/-sik- forms an " emphatic " (?) passive; e.g. **ut-** " to conquer " > **utsuk-** " to be conquered "; possibly a compound of **-sı:-/-si:-** and **-k-**; rare and unproductive.

-Vl- passive; common at all times and still productive.

-Vn- reflexive; turns transitive verbs into intransitive, and in some contexts forms passive (especially with verbs ending in **-a:-/-e:-**); common at all times and still productive.

(**-Vmsın-/Vmsin-**. See **-sı:-/-si:-**.)

-Vr- causative, possibly the oldest causative suffix (*ATG* 161); unproductive, may have some connection with **-z-**, q.v.; also a denominal suffix.

-dur-/-dür-/-tur-/-tür- the normal causative suffix of verbs ending in consonants; common at all times and still productive.

-ğur-/-gür- causative; it is not clear in what circumstances this suffix was used; it was attached only to verbs ending in certain letters including **-ç-**, **-d̲-**, **-r-**, **-ş-** and **-z-** but was not the only causative suffix attached to such verbs; productive perhaps until the mediaeval period.

-Vş- usually indicates action by more than one person, either reciprocal, e.g. **ur-** " to strike " > **uruş-** " to fight one another," or simultaneous, e.g. **igle:-** " to be ill " > **igleş-** " to be ill together," or competitive, with one person as the subject of the verb and the other an indirect object followed by **birle:**, or co-operative, with one person as the subject of the verb and the other as an indirect object in the dative; occasionally used with a single subject to indicate that the whole of it is involved in something, e.g. **boyma:-** " to be tangled " > **yip boymaşdı:** " the cord was completely tangled "; see Kāşğarī (*II*, 113 ff., 225 ff); common at all times and still productive.

-**Vz-** causative; rare and probably very old; may have some connection with -**Vr-**, cf. **-sız/-siz** and **-sıra:-/-sire:-**; unproductive.

-duz-/-düz-/-tuz-/-tüz- causative; rare and unproductive; probably old; described by Kāşğarī (*II*, 202) as specifically **Oğuz**; bears the same relation to **-dur-**, *etc.*, as **Vz-** to **Vr-**.

CHAPTER VIII

THE RECONSTRUCTION OF THE PHONETIC STRUCTURE OF PRE-EIGHTH CENTURY TURKISH

In Chapter V the evidence was assembled on which any attempt to reconstruct the phonetic structure of pre-eighth century Turkish must be based. Before proceeding to the actual work of reconstruction it may be useful to summarize the particular contributions which each of the languages concerned can be expected to make.

In some respects the **l/r** languages preserve a more archaic aspect than any others, although in other respects even the earliest of them have undergone sound changes from which the earlier forms of standard Turkish were immune. In the east the only substantial remains of such a language, presumably Tavğac, are the lowest layer of Turkish loan words in Mongolian discussed in Chapter XI. In some of these words we can trace the primitive initials **ḏ-** and **ñ-** which became *d-* and *n-* in Mongolian, and had already become **y-** in standard Turkish by the eighth century. In others we can trace the initial voiced post-palatal plosive **g-**, which survived in Oğuz but was devoiced in Türkü and the other eastern standard languages, as in " behind " *gerü* <***gérü:**> Uyğur **kérü:** ~ Osmanli **geri**. Whether this material will provide reliable evidence regarding the prevalence of initial **d-** in pre-eighth century Turkish is less certain. It includes some loan words with initial **d-**, like *dalay* " sea," Türkü **taluy** and *darğan* " freeman, privileged person," Türkü **tarxan,** but " camel " is *temeyen* and the Osmanli form **deve,** as opposed to Türkü **teve:,** suggests that this word may originally have had an initial **d-**. No evidence regarding an early Turkish initial **p-** can be expected from this material, since the sound did not exist in Mongolian.

The material from the western **l/r** languages provides evidence for the existence of initial **ḏ-** and **ñ-** but it is

doubtful whether it can be counted on to provide evidence regarding voiced and unvoiced initial plosives, unless these can be shown to have evolved into different initial sounds in Chuvash, which is not very probable.

The two earliest standard languages of which we have substantial remains, Türkü (including Manichaean Türkü) and Uyğur (including Uyğur-A), although they have lost initial **d̲-** and **ñ-**, have "levelled" the initial plosives, the labials to **b-**, the dentals to **t-** and the post-palatals to **k-**, and have also devoiced some of the old medial, and more particularly final, voiced plosives and the voiced affricate **c**, seem in most other respects to have retained a fairly archaic aspect, Türkü the more archaic of the two since it retained the old palatal **ñ** in the medial and final positions. The earliest remains of these languages also retain initial **b-** in words in which it subsequently became **m-** by regressive assimilation to a nasal later in the word.

Unfortunately we have very little direct evidence regarding the phonetic structure of Old Oğuz, the ancestor of the south-western group of languages including Osmanli/Republican Turkish, Türkmen and Azeri. Kāşğarī says that by the eleventh century it had already reached a more advanced stage of phonetic evolution than Xākānī in some respects, notably the elision of medial and final **ğ** and **g** in certain contexts, but it retained, and its descendants still retain, a wider repertoire of initial voiced and unvoiced plosives than Xākānī and also the medial and final voiced plosives and affricates which had been devoiced in that language, and the initial **b-** in words in which it had there become **m-**.

In the concluding chapter of his Introduction (*I*, 31-33) Kāşğarī made some interesting remarks on the phonetic differences between the dialects known to him, which are worth careful study, although they do not give direct evidence regarding the phonetic structure of pre-eighth century Turkish. Of all the early languages it is Xākānī that we know best owing to Kāşğarī's meticulous description of it. Its phonetic structure must have been very close

to that of Uyğur, and probably the differences between the two lay more in the field of vocabulary than that of phonetics.

With this preamble we can now proceed to try to establish what sounds existed in pre-eighth century Turkish. As every sound which existed in the language occurred in the medial position but some did not occur as initials and one not as a final, it will be best to start with the medial sounds. It is not my purpose to list all the words in which any sound, however rare, occurred; that is the business of lexicography. All that I want to do here is to establish what sounds occurred, and in what contexts; it will only occasionally be necessary to give examples of them.

VOWELS

There is no reasonable doubt that both long and short vowels existed in pre-eighth century Turkish. It is perhaps significant that both exist in some modern languages like Türkmen, but this would not by itself be conclusive. What is conclusive is the fact that all the three early alphabets enumerated in Chapter V which were equipped to distinguish between long and short vowels did so, and that, with some exceptions in the case of Runic, if a word or suffix was spelt with a long vowel in one of these alphabets it was also spelt with a long vowel in the others. The best evidence is that provided by Xākānī. Kāşğarī took great care to indicate long vowels in some words and suffixes and short vowels in others. Nearly all the Brāhmī texts made the same distinction, indicating long vowels in the same words and suffixes as those in which they were indicated in Xākānī. It also seems quite clear that in the Runic alphabet the vowel letter for **a/e** always, and the other vowel letters always except in the first syllable, represented long vowels in carefully written texts. The fact that, as shown above, there was a close association between long vowels and immediately following voiced plosives, and short vowels and immediately following unvoiced plosives shows that this was not a recent development but

must have been a characteristic of the language long before the eighth century.

The vowel sounds in pre-eighth century Turkish can therefore be tabulated as follows:—

Medial.	Back	Short	**a, ı, o, u.**
		Long	**a:, ı:, o:, u:.**
	Front	Short	**e, é, i, ö, ü.**
		Long	**e:, é:, i:, ö:, ü:.**

Three remarks only are required on this table.

(1) The existence of a closed **é** as well as an open **e** is not open to doubt. It is proved directly by the special letter for representing it in the Tuvan inscriptions, by the scriptions in the Brāhmī and Tibetan texts and by its existence in several modern languages. It is broadly true to say that if a word occurs with a closed **é** in one of these authorities it also occurs with a closed **é** in the others and *vice versa.* The manuscript of the *Nahcu'l-Farādīs* seems to be the only early text in Arabic script in which the scribe made a real effort to distinguish between **é, e** and **i** by writing the first *fatḥa yā*, the second *fatḥa* sometimes followed by *alif* and the third *kasra* sometimes followed by *yā.* The other early Arabic texts normally use the same scription, *kasra* sometimes followed by *yā*, to represent **é** as they used to represent **i**. This is, for example, the practice followed by Kāşğarī, but he does at any rate distinguish between **é/i** and **é:/i:**. In the other alphabets also, Sogdian, Uyğur and Manichaean Syriac, the same scriptions are used for **é** as for **i**. There does not seem to be any doubt that **é** occurred only in the first syllable of words. It occasionally appears in the second or a subsequent syllable in the Brāhmī texts, but this seems to be merely one of numerous aberrations of the scribes concerned. There does not seem to be any doubt that both **é** and **é:** occurred; this is not only *à priori* probable, but is also proved by the fact that in words in which the sound is known to have existed it is written in Kāşğarī with a *yā*, if long, as for example in **bé:rdi:** (*bā yā*) "he gave" (*III*, 180) and with a *kasra*, if

short, as for example in **bérildi:** (*bā maksūr*) " it was given " (*II*, 131).

(2) It is rather difficult to determine the extent to which vocalic harmony prevailed in pre-eighth century Turkish. So far as the distinction between back and front vowels is concerned, it is obvious that it was well established and that vowels of both kinds could not coexist within the same word, although there is some evidence in the Runic texts, more particularly the *Irk Bitig*, that some of the declensional and conjugational suffixes were regarded as containing front vowels, and that these front vowels survived even when such a suffix was attached to a word containing back vowels. So far as the question of unrounded and rounded vowels is concerned, the position is more obscure. There is some evidence that if the first vowel of a basic dissyllable was rounded the second also was rounded, at any rate in standard Turkish; the position may have been different in l/r Turkish, and this may represent the earlier practice. But it is very probable that suffixes like **-lığ/-lig** and **-sız/-siz** were originally regarded as containing unrounded vowels and that the practice of rounding them when they were attached to words containing rounded vowels was only beginning to establish itself in the eighth century.

(3) It is difficult to speak confidently about the quality of rounded vowels in the second and subsequent syllables. In some modern languages like Osmanli/Republican Turkish both **o** and **ö** and **u** and **ü** exist in the first syllables of words, but only **u** and **ü** in later syllables. The practice is different in some other modern languages like Kırğız. The Brāhmī and Tibetan alphabets are the only ones of those enumerated in Chapter V which distinguish between **o/ö** and **u/ü** and the scriptions in them show that Uyğur words containing **o** or **ö** in the first syllable normally contained the same vowel in the second syllable, and that **o** and **ö** might also occur in the second and subsequent syllables of words containing other vowels, perhaps only

a or **e**, in the first syllables. One curiosity of the spellings in the Brāhmī texts is **bo:** "this." The almost universal pronunciation to-day, except in the Lutsk dialect of Karaim (Radloff, *Opyt IV*, 1639), is **bu** and this is also the spelling in the *Codex Cumanicus* of the early fourteenth century, the next authority after the Brāhmī texts to distinguish between **o** and **u.** But the word occurs in twelve of the fifteen Brāhmī manuscripts, some thirty times in all, and is invariably spelt **bo:**. A spelling mistake therefore seems to be out of the question, but it is difficult to say whether this was the original pronunciation, and, if so, when it became **bu:**. In general, however, there is every reason to suppose that in the matter of these rounded vowels Uyğur retained the original vocalization of pre-eighth century Turkish.

Initial. Everything goes to show that the full range of eighteen (nine short and nine long) vowels could occur in the initial position.

Final. The scriptions in the Runic texts and Kāşğarī leave no room for doubt that the overwhelming majority of final vowels in Türkü and Uyğur were long. In Chapter V (1) (i) I collected the evidence which seems to prove that there was a short final vowel in the name **Türkü** and perhaps also in **Küli** and **şiri,** the latter in that event a loan word, but there is no other evidence of the existence of final short vowels in these languages. This is not of course conclusive proof that such vowels did not exist more widely in pre-eighth century Turkish. It will be shown in Chapter XI that a good many of the lowest layer of Turkish loan words in Mongolian carry final vowels which do not exist in any form of Turkish known to us. The balance of probability is that these were attached by the Mongols themselves, but it is conceivable that they still existed in Turkish when these words were borrowed and were elided before the eighth century except in the words mentioned above.

CONSONANTS

Medial. Medial consonants can occur either between two vowels (intervocalic), as the first member of a pair of consonants (CC), or as the second member of a CC other than a final CC. They may also occur as the first, second or third member of a CCC within a word; no word ends with CCC.

The following table shows all the consonants which seem to have existed in pre-eighth century Turkish.

	Plosive		Fricative		Nasal	Affricate		Sibilant		Semi-vowel
	v.	u.	v.	u.		v.	u.	v.	u.	
Labial	b	p	v	?	m	—	—	—	—	—
Dental	d	t	d̲	—	n	—	—	z	s	—
Denti-palatal	—	—	—	—	—	c	ç	—	—	—
Palatal	—	—	—	—	ñ	—	—	?	ş	y
Post-palatal	g	k	?	—	} ŋ	—	—	—	—	—
Velar	?	k	ğ	x		—	—	—	—	—

Liquids:—l, r.

So far as the intervocalic position is concerned, most of these consonants occurred frequently both within basic words and as the last letter of a basic word to which a suffix beginning with a vowel was attached, but something must be said about a few individual letters.

Apart from the rather incoherent scriptions in Brāhmī, the earliest authority who distinguished quite clearly between **d** and **d̲** was Kāşğarī. Intervocalic **-d-** occurred constantly in Xākānī as the first letter of such suffixes as **-dı:** attached to basic verbs ending in a vowel, but intervocalic **-d-** within a basic word of such a form as VdVC seems to have been very rare. In most words in which it appears in printed texts, e.g. *tduk* (for **ıd̲uk**) " sacred," it is an error for **-d̲-**. The only certain case of such an intervocalic **-d-** that I can recall is in **idiş** " cup, vessel," but there may be a few others. It must also have occurred in pre-eighth century Turkish, and still occurs in languages of the Oğuz group such as Osmanli/Republican Turkish, when a suffix beginning with a vowel is attached to a monosyllabic basic word ending in **-d**, e.g. **a:d** " name " >

adı: " his name," but such final **-d**'s were devoiced in Türkü, Uyğur and Xākānī.

Intervocalic voiced **-c-** occurred in languages of the Oğuz group, and so presumably also in the pre-eighth century Turkish, both within basic words like **acı:-** " to be bitter " and at the end of monosyllabic basic words which might be followed by a suffix beginning with a vowel like **a:c** " hungry." In the second case the **-c** was almost certainly devoiced in Türkü *etc.*, whether it was in the first case we do not know.

As stated in Chapter V (1) (i) **ñ** was always a rare sound but certainly existed in the intervocalic position.

It was pointed out in Chapter V (1) (ii) that there is some evidence in the Brāhmī texts that the voiced post-palatal **g** was sometimes a fricative rather than a plosive, at any rate in an intervocalic or final position; conversely the voiced velar **ğ** may sometimes, for example in CC clusters, have been a plosive rather than a fricative.

There is no direct evidence that **x** ever occurred in the intervocalic position in the early period. Certainly it never did in Kāşğarī, the first authority to distinguish clearly between **x** and **k**.

It will be seen that three sounds in the table in Chapter III do not appear in this table, and these omissions should perhaps be explained. The unvoiced labial fricative **f** appears in an intervocalic position in our manuscript of Kāşğarī, but the language in *I*, 8, where it is described as *al-fā'u'l-'arabīya*, and the pronunciation of such words in other Turkish languages suggest that these spellings are scribal errors. It also seems to have occurred in the tribal name **Afar** (for Avar), spelt *pe resh* in the Runic texts. The question whether it occurred as the first member of a CC is discussed below.

Kāşğarī says specifically (*I*, 31) that the Oğuz pronounced **v** as **w**, implying that the latter was a secondary sound not native to the language. This is no doubt the case, although a few words, presumably dialect forms, are spelt with **-w-** in our manuscript of Kāşğarī.

The voiced palatal sibilant **j** was a common sound in Sogdian, and occurs in one or two Sogdian loan words in Xākānī like **erej** " happiness, bliss " (*KB* 432 *etc.*), representing Sogdian **ré:j** with a prosthetic **e-**. It also occurs in a few other words listed by Kāşğarī like **üjme:** " mulberry." So far as I know, it has not yet been proved that these too are Sogdian loan words, but it is probable that they are and that **j** was never a native Turkish sound.

Most, but not all, of the sounds in the table could occur either as the first or as the second member of a CC, but the number of other sounds with which each of them could be associated in this way, except when the first C was the last sound of a basic word and the second the first sound of a suffix, was strictly limited. Broadly speaking, it would be true to say that **r**, and to a less extent **l**, are the only sounds which associate really freely with other sounds as the first member of a CC and that **r**, and to a less extent **d**, **t** and **k**, are the only sounds which associate freely with other sounds as the second member. The number of possible combinations in a final CC was very limited indeed. The table on the next page shows all the combinations which I have found in Türkü, Uyğur and Xākānī. Some of them are very rare, and the table may not be quite complete, as I may have missed one or two other rare combinations. I have tried to exclude all combinations which occur only when the first C is the last sound in a basic word and the second the first sound in a suffix, but I may not always have succeeded. Combinations which could occur at the end of a word are marked with an asterisk.

It is possible that some of the very rare combinations occur only in onomatopoeics (for example I have noted **pg** only in **üpgük** " hoopoe," which looks like an onomatopoeic), in words carrying an obsolete suffix which I have failed to identify, or in words which seem to be basic but are in fact suffixed words derived from obsolete dissyllables of such forms as VCVC or CVCVC in which the second V has been elided when a suffix beginning with a V was attached to it.

		b	p	v	ç	d	t	ğ	k	g	k	l	m	n	ŋ	ñ	r	s	ş	z
	b	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	br	—	—	—
	p	—	—	—	—	—	—	—	—	pg	pk	—	—	—	—	—	pr	—	—	—
	v	—	—	—	vç	vd	—	—	—	—	—	—	—	—	—	—	vr	vs	—	vz
	ç	—	—	—	—	—	—	çğ	çk	—	çk	—	—	—	—	—	—	—	—	—
	d	—	—	—	—	—	—	dğ	—	—	—	—	—	—	—	—	dr	—	—	—
	ğ	—	—	—	—	—	ğt?	—	—	—	—	ğl	ğm	ğn	—	—	ğr	—	—	—
velar	k	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	kr	—	kş	—
	x	—	—	—	—	—	xt?	—	—	—	—	—	—	—	—	—	—	xs	xş	—
	g	—	—	—	—	gd	—	—	—	—	—	gl	gm	—	—	—	gr	gs	—	—
post-palatal	k	—	—	—	kç	—	kt	—	—	—	—	—	—	—	—	—	—	—	kş	—
	l	—	lp*	—	lç	ld	lt*	—	lk*	lg	lk*	—	—	—	—	—	—	—	—	—
	m	—	—	—	—	—	—	mğ	—	mg	—	—	—	—	—	—	mr	—	—	mz
	n	—	—	—	nç*	nd*	nt*	—	—	ng?	—	—	—	—	—	—	—	—	—	—
	ŋ	—	—	—	—	ŋd	—	—	—	—	—	ŋl	—	—	—	—	ŋr	—	—	—
	r	rb	rp*	rv	rç*	rd	rt*	rğ	rk*	—	rk*	—	—	—	rŋ	rñ	—	rs*	—	—
	s	—	—	—	—	—	st*	—	—	—	sk*	—	—	—	—	—	—	—	—	—
	ş	—	—	—	—	—	—	—	—	şg	şk	—	—	—	—	—	—	—	—	—
	y	—	—	—	—	—	—	—	yk	—	—	—	ym	yn	—	—	yr	—	—	—
	z	—	—	—	—	—	—	—	—	zg	—	—	—	—	—	—	—	—	—	—

The combinations **xs** and **xş** are attested in Xākānī **axsa:-** " to limp " and **oxşa:-** " to resemble "; neither of these words occur in Türkü, the second occurs in Uyğur, but there is no means of discovering whether the combination in that language was **-xş-** or **-kş-**, and it is possible that **-xs-** and **-xş-** are secondary forms of **-ks-** and **-kş-** or **-ğs-** and **-ğş-**. Conversely **-kş-** may be a secondary form of **-xş-**.

As regards **-ğt-** and **-xt-**, Kāşğarī normally spells " to roll over " **axtar-**, but says (*I*, 219) that the **x** was " altered from **ğ**."

So far as final CC are concerned it will be seen that the only known first members are **l, n, r** and **s**. Of these **s** is very rare; I know of it only in **ast** " below " and **üsk** " in the presence of," neither of which normally occurs without a suffix.

There is some doubt whether both **-nd** and **-nt** occurred simultaneously as final CC's. Kāşğarī, for example, in *III*, 7 spells the generic name for " horse " **yond,** but in Türkü (*IB*) it is spelt **yont**. It is likely that one, or the other, perhaps more probably **-nd,** is the original pre-eighth century form.

It is possible that **f,** although it is not a native Turkish sound in the initial, intervocalic or final position, may have

existed, presumably as a devoiced secondary form of **v**, as the first member of a CC. There are one or two words in our manuscript of Kāşğarī like **yufğa:** " soft " (the **-ğa:** is probably a suffix) which contain **-fğ-**, but these may well be scribal errors; **yafğu:**, for example, must surely be an error for **yavğu:**.

Initial. The range of initial sounds was more restricted, although not as restricted as is generally stated. The position seems to have been as follows.

Labials. I set out in *The Initial labial sounds in the Turkish languages* the evidence, taken principally from the Oğuz group, which seems to prove that both **b-** and **p-** were initial sounds in pre-eighth century Turkish. In the same paper I pointed out that initial **v-** occurs in some modern languages of the Oğuz group in three basic words **var** " existing," **var-** " to go " and **vér-** " to give " and in words derived from them. If this was merely a secondary form of **b-**, it is difficult to see why it should have evolved only in these words and not in other very similar words like **bi:r** " one "; but initial **v-** was unknown to Kāşğarī even in Old Oğuz, and the sound cannot be traced further back than the thirteenth century.

Affricates. Initial **c-** is fairly common in some modern languages but only as a secondary form of **y-**; it seems probable that in pre-eighth century Turkish the only initial affricate was **ç-**.

Dentals. There is a good deal of evidence, derived not only from the Oğuz group, that although in Türkü, Uyğur and Xākānī the only initial dental plosive was **t-**, initial **d-** also existed in pre-eighth century Turkish. For example, although " to say " was in those languages **té:-**, it is **dé-** in all modern languages except those, like Khakas, which represent even initial **b-** by **p-**.

Although by the eighth century initial **d̲-** had become **y-** in standard Turkish, there is conclusive evidence from the remains of the **l/r** languages both in the east and in the west that it existed in pre-eighth century Turkish.

Velars. Initial **k-** is one of the commonest initials in the language, but there is no evidence that there was ever an initial **ğ-**. Initial **x-** occurs in a few words listed by Kāşğarī, the first authority in which this sound can be read with certainty. In most cases it is specifically said to be a secondary form of **k-**; some of the other words concerned are loan words. The one word in which it certainly occurred as an original initial is **xa:n** " king " (*III*, 157). This is the kind of word which might have survived from an earlier period; and taking this with the fact that the Chinese and other foreign authorities consistently spelt the tribal name Hun with an initial **x-** (sometimes represented by *h-*), it seems reasonably clear that an initial **x-** existed in pre-eighth century Turkish, probably as a rare sound which had almost always evolved into **k-** by the eighth century.

Post-palatals. In Türkü, Uyğur and Xākānī the only post-palatal initial was **k-**, but there is sufficient evidence in modern languages, mainly in the Oğuz group, and in a few early loan words in Mongolian to show that initial **g-** also existed in pre-eighth century Turkish.

Nasals. As I showed in *The initial labial sounds in the Turkish languages*, initial **m-** occurs only in loan words and as a secondary form of **b-**, when this sound has been nasalized by regressive assimilation to a nasal later in the word. Initial **n-** occurred only in the basic words **ne:** " what? " and **ne:ŋ** " thing " *etc.* and in words derived from them. Although by the eighth century initial **ñ-** had become **y-** in standard Turkish, there is conclusive evidence from the remains of the **l/r** languages both in the east and in the west that it existed in pre-eighth century Turkish. Initial **ŋ-** never existed in Turkish even in loan words.

Sibilants. There is no doubt that there was always an initial **s-** in Turkish, although it is not very common. The existence of a genuine Turkish initial **ş-** is less certain. It occurred in loan words from Türkü onwards. The evidence from texts in Uyğur script is unsatisfactory since it is seldom possible to distinguish definitely between **s-** and **ş-**. There

are about forty words beginning with **ş-** in Kāşğarī. In some it seems to be a secondary form of **s-**, and in others of **ç-**, under the influence of an **-ş-** later in the word. Some of the rest are clearly loan words. Only very careful word-by-word examination would make it possible to determine whether there is a residuum of words with a genuine initial **ş-** in Xākānī. Whatever the conclusion is, it would probably be equally applicable to pre-eighth century Turkish. It seems clear that there was never an initial **z-** in Turkish. Loan words with initial **z-** began to enter Turkish from the Manichaean period onwards, but it appears that in the early period the sound was felt to be so alien to the language that a prosthetic vowel was often put before it, e.g. **ezrwa:** < Sogdian *Zurwa:*.

Liquids. There is good reason to suppose that there was never an initial **r-** in Turkish. There are no words with this initial in Türkü, Uyğur or Xākānī except obvious loan words, and although such loan words normally appear in their original form in religious texts, it is clear that in the early period the sound was felt to be so alien to the language that in popular speech a prosthetic vowel was put before it, for example, **erej** "happiness, bliss" < Sogdian **ré:j** (*KB* 432 *etc.*) and **ertini** the Uyğur and Xākānī scription of Sanskrit *ratna* "jewęl," in Sogdian *ratni*. The question of initial **l-** is less certain. It is generally believed that there was never such a sound in the language, but there is no sign that it was ever felt to be so alien that a prosthetic vowel should be placed before it, and in fact, apart from a number of words which are certainly or almost certainly loan words, there are several words with initial **l-** in the early languages:—

la:çın "falcon," in Uyğur, Xākānī and several later languages;

lağzın "pig," in Türkü and Uyğur;

lorzı: (or **lurzı:**?) "club, mace," in Uyğur;

lıyu: (or **liyü:**) "mud," in Xākānī

and perhaps one or two more. It is probable that these are

loan words of which the origin (perhaps "Tokharian") has not yet been discovered, but until this has been proved, it is impossible to say categorically that there was not an initial l- in pre-eighth century Turkish, though obviously, if there was, it was a very rare sound.

Thus the following table of initial sounds can be constructed:—

	Plosive		Fricative		Nasal	Affricate	Sibilant	Semi-vowel
	v.	u.	v.	u.		u.	u.	
Labial	b-	p-	—	—	—	—	—	—
Dental	d-	t-	d̲-	—	n-	—	s-	—
Denti-palatal	—	—	—	—	—	ç-	—	—
Palatal	—	—	—	—	ñ-	—	? ş-	y-
Post-palatal	g-	k-	—	—	—	—	—	—
Velar	—	k-	—	? x-	—	—	—	—

Liquids:- ?? l-.

Final. The only medial consonantal sound which did not occur as a final in any of the early languages is **x**. All the rest occurred at the end of basic monosyllabic nouns and verbs, there being an affinity, as indicated above, between some of these sounds and preceding long and short vowels respectively.

So far as occurrences at the end of longer words is concerned, however, the position is a good deal more complicated.

The sounds fall into seven distinct classes:—

(1) Four voiced sounds **-b, -v, -c** and **-d** do not seem to have occurred at the end of words longer than monosyllables.

(2) The sound **-p** does not seem to have occurred at the end of any words longer than monosyllables except as the gerund suffix -V**p**.

(3) The sound **-ñ** occurred at the end of one or two dissyllabic basic nouns.

(4) The sound **-s** occurred at the end of a few dissyllabic basic nouns and even fewer verbs; it did not

occur at the end of suffixes (the negative suffix **-ma:s/-me:s** in Xākānī is a later form of **-ma:z/-me:z**).

(5) Six sounds, **-ç, -ğ, -g, -m, -ŋ** and **-y,** occurred at the end of basic dissyllabic nouns, and as, or at the end of, nominal suffixes; **-m** and **-ŋ** were also possessive suffixes and **-y** occurred at the end of verbal suffixes.

(6) Four sounds, **-t, -l, -ş** and **-z** occurred at the end of dissyllabic basic nouns, as nominal and verbal suffixes, and at the end of nominal and conjugational suffixes.

(7) Five sounds, **-ḏ,** velar and post-palatal **-k, -n** and **-r,** occurred at the end of dissyllabic basic nouns, and a few dissyllabic verbs which seem to be basic, but as all these sounds were more or less common verbal suffixes, it cannot be taken as certain that these verbs are basic and not derived forms of obsolete basic verbs. All these sounds except **-ḏ** also occurred at the end of suffixes, including some declensional and conjugational suffixes.

To sum up, dissyllabic basic nouns could end in **-ç, -t, -ḏ, -ğ,** velar **-k, -g,** post-palatal **-k, -l, -m, -n, -ŋ, -ñ, -r, -s, -ş, -y** and **-z**; no dissyllabic basic verb could end in any consonant except **-s-** and possibly **-ḏ-,** velar and post-palatal **-k-, -n-** and **-r-,** but there are very few dissyllabic basic verbs with consonantal endings.

No basic words except monosyllables could end in two consecutive consonants; the only sounds which could occur at the end of such words were **-p, -ç, -d** or **-t** (or perhaps both) and velar and post-palatal **-k** and only a very limited number of sounds could precede each of these final sounds. The only longer words ending in two consecutive consonants were those carrying the nominal suffix **-ç,** preceded by **-n-**.

CHAPTER IX

THE LATER HISTORY OF THE UYĞUR ALPHABET

In Chapter V (1) (iv) I discussed the use of the Uyğur alphabet as a medium for writing native Turkish words. The history of that alphabet was, therefore, broken off at a point when it was just entering upon a phase of great interest which has never, so far as I know, been systematically studied.

When the Uyğur language first began to be written in Uyğur script, it already contained some loan words, and when it became a medium for preaching Buddhism it acquired a good many more. Some of these loan words contained sounds which were foreign to Turkish, and means had to be devised for writing them with Uyğur letters.

No problems arose regarding writing Sogdian loan words; they were normally taken over in their original spelling. But problems did arise when the alphabet had to be used to represent sounds in loan words from other languages. However, here too, so far as loan words, including proper names, from Sanskrit and other Indian languages and Chinese were concerned, Sogdian had already paved the way, and, generally speaking, the Sogdian practice was followed. Probably the only serious problem was the representation of the simple aspirate *h*. In Sogdian this sound, at any rate in association with back vowels, was regarded as a variant of *x* and was represented by *gimel-cheth*, and this scription was adopted in Uyğur not only for writing Sanskrit names like *Harichandra* but also for writing the name of Herod (**Hırodıs**) in the Christian apocryphon in F. W. K. Müller, *Uigurica*, A.K.P.A.W., Berlin, 1908. But this scription was not appropriate in words with front vowels, and in them *h* was represented by *caph*.*

* See von Gabain, *Die Uigurische Übersetzung der Biographie Hüen-tsangs*, I, S.P.A.W., Berlin, 1935, p. 21, note 22.

The problem became more acute in the tenth century or even earlier, when the Turks entered into friendly contact with the Arabs and many of them were converted to Islam. The alphabet now had to be adapted to represent a whole range of Arabic sounds foreign to Turkish. Fortunately we have an almost contemporary witness to tell how it was done. Kāşğarī did so in the passage in his *Dīwān*, translated in the Appendix to Chapter V, though he hardly realized the full extent of the problem, and had difficulty in distinguishing between the concepts of "letter" and "sound." It will be seen that he considered that the Uyğur alphabet contained eighteen letters including hooked *resh* of which three, *cheth*,* *nun* and *schin* had to be identified by dots. These eighteen letters had basic phonetic values, which he sets out, and seven when differentiated by a single dot (*nuqṭa*) had additional phonetic values which were also Turkish sounds:—**v/f, x/ğ, z/j, k/g, b/p, ç/c,** and "nasal caph" used to represent **ŋ**. As I pointed out in Chapter VIII, it is doubtful whether **f** and **j** were really Turkish sounds. It is also doubtful whether **x, ğ** and **k** were ever really written *gimel*, *gimel* with one dot, *gimel* with two dots. The last entry is a muddle; **ŋ** was written *nun caph*. In order to represent purely Arabic sounds he said that three—*ṯ*, *ḏ* and *ṣ*—were represented by Uyğur letters with dots and three—'(*'ayn*), *h* and *ḥ*—by *gimel* differentiated by a mark (*'alāma*), that is no doubt a miniature Arabic letter. It will be seen that he says nothing about the use of *caph* to represent *h*, presumably because he regarded all Arabic words as containing back vowels, so that for that purpose only *gimel* would have been appropriate. It will be seen that there are certain omissions. Apart from the omission of **d**, which was mentioned in Chapter V (1) (iv), nothing is said about **ṭ** and **ḍ**, which must have been represented by appropriate letters differentiated by dots or, more probably, "marks," or about *hamza*. Initial *hamza*

* It is doubtful whether he really understood the relationship between *gimel* and *cheth*, which had identical outlines, but apparently he regarded the first as representing an affricate, transcribed *xā*, and the second, with two superscribed dots, as representing a plosive, transcribed *qā*.

was no doubt regarded as the same as a smooth vocalic ingress and represented by *aleph*, and medial *hamza* was probably treated in the same way as *'ayn* and represented by *gimel-cheth* probably with a " mark." The scriptions in the fifteenth century manuscripts mentioned below also suggest that in the initial position *'ayn*, **h** and perhaps even **h** were represented not by *gimel-cheth* but by *aleph* with " marks."

The following table, therefore, shows the full range of phonetic values of the Uyğur letters, when adapted in the way described by Kāşğarī, with the adjustments suggested above, for writing Arabic and other loan words in Turkish, values which should properly be identified or differentiated by dots (*nuqṭa*) being accompanied by (*n*), and those which should be differentiated by " marks " (*'alāma*) being accompanied by (*m*):—

aleph	**a, e;** as an initial (1) smooth vocalic ingress; (2) *hamza* (m.?); (3) '- (*m.*); (4) **h-** (*m.*); (5) probably **ḥ-** (*m.*); see *zain* and *nun*.
beth	**v; f** (*n.*).
gimel-cheth	**x; ğ** (*n.*?); velar **k** (2 *n.*); medially also (1) -'- (*m.*); (2) **-h-** (*m.*); (3) **-ḥ-** (*m.*); (4) probably *hamza* (*m.*).
vau	**o, u; ö, ü; w.**
zain	**z; j** (*n.*); **ẓ** (*m.*); sometimes used for *samech-schin*; as a final indistinguishable from *aleph*, and *nun*, if undotted.
jod	**é, ı, i; y;** used to modify *vau*.
caph	**g,** post-palatal **k;** possibly medial *hamza* (*m.*) and **h** (*m.*) with front vowels.
lamed	**d; ḏ; ḍ** (*n.*); **ṯ** (*n.*); sometimes used for *tau*.
mem	**m.**
nun	**n** (*n.*); *nun caph* **ŋ**; *nun jod* **ñ**; if undotted, indistinguishable from *aleph* and final *zain*.
samech-schin	**s; ş** (2 *n.*); **ṣ** (*n.*); sometimes used for *zain*.
pe	**b; p** (*n.*).

tzaddi ç; c (*n.*).
resh r.
tau t; ṭ (*m.*?); sometimes used for *lamed.*
hooked *resh* l.

This is the script which according to Kāşğarī was used for Turkish diplomatic correspondence between Kāşğar and China and in other parts of the Turkish-speaking world; it was also no doubt used for ordinary day-to-day affairs by the pagan Turks, but there is no evidence that it was used to write Turkish west of the Pamirs before the Mongol invasion. Indeed there is no reasonable doubt that when the Karakhanid Turks were converted to Islam the Arabic alphabet became the ordinary alphabet for writing Turkish in the Moslem world. Ordinary commercial documents of the Karakhanid period written in Arabic script have been discovered in Sinkiang, and there is no doubt that literary works like the *Kutadgu: Bilig* were written in the Arabic script. It was the Mongols who carried the Mongol Official Alphabet, which was in effect the Uyğur alphabet, across the Pamirs and made it the official alphabet in Transoxiana, Persia and their other western dominions. Originally used in these areas only to write Mongolian, it remained the official alphabet when the official language became Turkish. It was in this alphabet, and not in an Uyğur alphabet handed down continuously from generation to generation by Turkish scribes, that the fifteenth century Turkish manuscripts in " Uyğur " script were written.

Unfortunately no Uyğur text between the Turkish conversion to Islam and the Mongolian conquest which uses the system of orthography described above with its elaborate system of dots and " marks " has survived. Two or three texts, for example documents Nos. 9 and 88 in Radloff's *Uigurische Denkmäler*, are known which contain Arabic and Persian loan words and so have an Islamic background, but they are almost completely lacking even in dots. To see how the system actually worked we must turn to the fifteenth century manuscripts just mentioned.

The story of the origin of the first Mongolian Official Alphabet (hereafter referred to as M.O.A.) is well known*; and there is no reason to suppose that it is not authentic, at any rate in broad outline. It is that when Chinggis Xan finally routed the Naiman in A.D. 1204 he captured the Naiman Xan's Chancellor, an educated Uyğur called, in Chinese transcription, *T'a-t'a T'ung-a* (probably **Tatar Toŋa:**) and his official seal. Enquiring about the nature of the latter, he was informed of the utility of writing for the conduct of official business, a subject of which he had previously been completely ignorant, and ordered T'a-t'a T'ung-a to teach his four sons to write the Mongolian language in the Uyğur script. At this time the Uyğur alphabet had been in use for well over a century to write Turkish containing Arabic and other loan words, and there is no doubt that with his diplomatic background he knew perfectly well how to use it to write sounds which were foreign to Turkish. There were in fact only two such Mongolian sounds, *h* and the intervocalic hiatus, and the alphabet just described was perfectly suitable for writing both of them. It is obvious that T'a-t'a T'ung-a himself wrote a good hand; certainly the earliest Mongolian inscriptions and manuscripts would, if they had been Turkish, have been described as being in a hand much superior to that of the contemporary or later Turkish documents and texts. There is, however, no trace in any Mongolian texts of " marks "; it would obviously have been unsuitable to use miniature Arabic letters in them.

As will be shown in Chapter X, the phonetic range of Mongolian was more restricted than that of Turkish, and the phonetic values of the letters of the Uyğur alphabet as we may assume that it was used by T'a-t'a T'ung-a for writing Mongolian were as follows:—

aleph	*a*, *e*; as an initial, smooth vocalic ingress and *h*.
gimel-cheth	*ğ*, velar *k*; intervocalic hiatus with back vowels.

* See for example R. Grousset, *L' Empire Mongol* (*1re phase*), Paris, 1941, p. 267.

vau	*o, u; ö, ü.*
zain	(used only at the end of words) *s.*
jod	*é, i; y;* also used to modify *vau.*
caph	*g,* post-palatal *k;* intervocalic hiatus and *h* with front vowels.
lamed and *tau*	*d, t.*
mem	*m.*
nun	*n* (*n.*).
samech-schin	*s, ş* (2 *n.*).
pe	*b.*
tzaddi	*ç, c.*
resh	*r.*
(*tau*, see *lamed.*)	
hooked *resh*	*l.*

It is doubtful whether the two dots over *gimel* were used to differentiate velar *k* from *ğ*, but they were certainly used to identify the letter and distinguish it from two consecutive *alephs*/(undotted) *nun's*. *Beth* (*v*/*f*) was not required to write Mongolian, but was used in loan words in certain circumstances mentioned below.

A description of this alphabet dating from the third quarter of the fourteenth century has been preserved in a manuscript* of a *mecmua* or miscellany, which was compiled for a Turkish notable living in Persia. It is headed *fī xaṭṭi'l Uyğur al-ān yuqāl Muğūlī*, "on the Uyğur script now called Mongolian."† It states that the alphabet contained seventeen letters, though in fact eighteen plus three variant forms are enumerated, each with a transcription in Arabic. The forms of the letters are much the same

* Now in the library of the Ministry of Education at Ankara. A facsimile of the relevant part was published in R. R. Arat, *Atebetü'l-Hakayik*, İstanbul, 1951, pp. clxix ff. See also pp. 34 ff. of that work, and the same author's *Kutadgu: Bilig*, Vol. I, Istanbul, 1947, pp. xxxix ff. A smaller part of the text had been published earlier by the same author in an article in *Muallim M. Cevdet*, a miscellany compiled by O. Ergin, İstanbul, 1937, p. 673. This article also contains reproductions of other later texts of the same kind.

† This description is not an isolated one. One of the manuscripts of the *Atebetü'l-Hakayik*, in the same script with an interlinear transcription in Arabic script, dated A.D. 1480, is described on the fly-leaf as "a miscellany in the Mongolian language" (which is, of course, false) and *bi'l-xaṭṭi'l-Muğūlī*, *op. cit.*, p. lxiii.

as those in the manuscript of Kāşğarī and the transcriptions of individual letters too are the same, with the following differences from those set out in the Appendix to Chapter V:—

(a) the third letter (*gimel*) carries two dots and is transcribed *qā*, while the sixth (*cheth*) written in final form and also carrying two dots is also transcribed *qā*.

(b) the fifth letter has the form of *samech* and is transcribed *sā*, while the twelfth has the form of a final *zain* and is transcribed *zā*.

(c) the ninth letter has the form of *tau* and is transcribed *tā*, while the seventeenth has the form of *lamed* and is transcribed *ṭā* (probably intended to represent *dā*).

(*d*) *pe* carries a superscribed dot and is transcribed *pā*.

The three variant forms which follow hooked *resh* are *samech* transcribed *şā*, *mem* transcribed *mā* and *gimel-cheth* with two superscribed dots transcribed *qā*, all in final form. The practice of adding three letters in final form goes back to the earliest alphabet known, that in a Manichaean manuscript mentioned in Chapter V (1) (iv), and some of the changes mentioned must go back two or three centuries.

The number of surviving Mongolian inscriptions and documents of the thirteenth and fourteenth centuries written in this script is small, and the number of those of which facsimiles have been published is smaller still. B. Ya. Vladimirtsov in his *Sravnitelnaya Grammatika Mongolskogo Pis'mennogo Yazyka i Khalkhaskogo Narechiya*, Leningrad, 1929, p. 34, published a list of those then known. Others have since been published by Prof. F. W. Cleaves in a series of articles in the Harvard Journal of Asiatic Studies from 1949 onwards, and one of these also by Prof. E. Haenisch in *Mongolica der Turfan Sammlung*, A.D.A.W., Berlin, 1954. The latest publication of such facsimiles is in Prof. Rinchen's article *L'inscription sino-mongole de la stèle en l'honneur de Möngke Qayan*, Central Asiatic Journal IV, 2, 1959.

Careful examination of these facsimiles shows that the script used is that of the Mongolian Official Alphabet described above and in particular:—

(1) that the only diacritical points used, and those only sporadically, are:—

(a) two dots over *gimel-cheth* to identify that letter, without reference to its phonetic value;

(b) a single superscribed dot identifying *nūn*; and

(c) two subscribed dots identifying *schin*;

(2) that *zain* hardly ever occurs, and then only at the end of a word to represent *s*.

(3) that *lamed* and *tau* are used indiscriminately to represent both *d* and *t*.

As has been explained above, the M.O.A. had originally been devised for purely civil and governmental purposes while the Mongols were still pagans. When Turkish missionaries began to convert the Mongols to Buddhism they naturally used their own Uyğur alphabet to write Mongolian, and retained the Uyğur orthography for the numerous Turkish loan words (including some of foreign origin) which they introduced into the language. They retained the old method of representing non-Turkish sounds and had probably never heard of the use of "marks" to differentiate some of them. The officials, however, T'a-t'a T'ung-a and his colleagues and successors, were more concerned with the use of the alphabet, developed as described by Kāşğarī, as a medium for diplomatic correspondence. Thus right from the beginning the civil and the Buddhist religious alphabets went their own ways, and very soon at any rate on one point, the representation of some labial sounds, quite different practices developed. For reasons which are well known the Mongolian rulers of Persia wished to establish diplomatic relations with France in the last quarter of the thirteenth century, and letters addressed by Arğun to Philippe le Bel in A.D. 1289 and by Ölceytü to that same monarch in A.D. 1305 have survived, see Prince Roland Bonaparte, *Documents de*

l'époque mongole, Paris, 1895, plate 14. The word "France" obviously presented a difficulty, certainly of spelling and perhaps also of pronunciation. In the earlier letter it is spelt with an initial *pe*, *Baransa*, *Paransa* or *Faransa*? In the latter it is spelt with initial *beth Varans* or *Farans*?

That, so far as I have been able to discover, is the last, perhaps the only, appearance of *beth* west of the Pamirs. Mongolian documents from this area are extremely rare. The only one in M.O.A. that I have been able to trace is the silver *pai-tzü** issued by a certain 'Abdullah of the Golden Horde, which bears his name spelt *aleph*, *aleph*, *pe*, *lamed*, *vau*, hooked *resh*, hooked *resh*, *aleph*. It will be seen that the initial *'ayn* is represented by *aleph* and that the final *hā* is not represented by anything. All the other documents from the west of the Pamirs in the M.O.A. are in Turkish. It was only natural that when Turkish became the language of culture in the Mongolian (and quasi-Mongolian) courts of Persia and Transoxiana, a demand arose for copies of the Turkish classics in the official script, and for such works with their plethora of Arabic and Persian loan words the full apparatus of dots and "marks" was essential if they were to be made intelligible. In the Vienna manuscript of the *Kutadğu: Bilig*, of which Radloff published a facsimile in *Kudatku Bilik*, St. Petersburg, 1890 (of which a reduced facsimile was published in İstanbul in 1942), we have the first draft, in an abominable hand with numerous erasures, of a transcription of this work in the M.O.A., from which a calligraphic *edition de luxe* was to be prepared. Other manuscripts in the same script, like the Paris manuscript Bibliothèque Nationale, Supplément Turc, No. 190,† dated A.H. 840 (A.D. 1436-37), the British Museum manuscript Or. 8193,‡ dated A.H. 835 (A.D. 1431), and one of the manuscripts of the *Atebetü'l-Hakayik* reproduced in Arat

* Found near the Dniepr in A.D. 1845 and since reproduced more than once, for example in *The book of Ser Marco Polo*, edited by Sir Henry Yule, 3rd edition revised by H. Cordier, London, 1903; *I*, p. 355.

† Large sections were reproduced in facsimile in Pavet de Courteille's two books, *Mirâdj Name*, Paris, 1882, and *Tezkereh-i Evliya*, Paris, 1889.

‡ Described in my article *A hitherto unknown Turkish manuscript in "Uighur" characters*, J.R.A.S., 1928, with two reproductions.

op. cit., are the final and sumptuous calligraphic texts. Most of these texts are in Çağatay, a language in which the old Turkish **v** had become **w**, and the two Xākānī texts, the *KB* and the *Atebetü'l-Hakayik*, have been transposed into Çağatay spelling with **v** represented by *vau*.

One manuscript stands in a class by itself, the Paris manuscript Bibliothèque Nationale Supplément Turc, No. 1001, commonly called the *Oğuz Name*; Radloff's *Kudatku Bilik* contains a facsimile of some pages of it, and the best edition of the text is W. Bang and G. R. Rachmati (R. R. Arat), *Die Legende von Oghuz Qaghan*, S.P.A.W., Berlin, 1932. The language of the text seems to be an Oğuz dialect with a number of Mongolian loan words. The script is a very distinctive and illegible one with quite a number of dots but no marks, the latter perhaps because it contains few, if any, Arabic loan words; but fundamentally it is an M.O.A. script with **v** (or **w**?) represented by *vau*.

In principle the alphabet used in these manuscripts is the Uyğur alphabet adopted as the M.O.A. with the full apparatus of dots and marks prescribed by Kāşğarī, but modified in the following respects:—

(1) *aleph* is used to represent initial **'**, **h** and **ḥ**, theoretically differentiated by the appropriate "marks" when so used.

(2) *beth* does not occur.

(3) *gimel-cheth* when representing **ğ**, velar **k** or **x** is often differentiated by a "mark."

(4) both *zain* and *samech-schin* are used to represent **ḏ** (*m.*), **ḍ** (*m.*), **ṯ** (*m.*), **s**, **ṣ** (*m.*), **ş** (*n.*), **j** (*n.*), **z** and **ẓ** (*m.*).

(5) *caph* represents **g**, post-palatal **k** and sometimes apparently also medial *hamza* (*m*?) and **h** (*m*?).

(6) *lamed* and *tau* both represent **d**, **ḏ**, **t** and **ṭ** and are often differentiated by "marks" with all these values.

(7) *pe* represents **f** as well as **b** and **p**.

It must be emphasized that the "marks" are often omitted, and that it is sometimes quite impossible to

determine what word is intended, particularly if it is an Arabic loan word.

Thus in the west the official variety of the M.O.A. lingered on until the fifteenth century as a kind of formal script for use at court, just as Gothic black letter survived in England as a formal script long after it had been displaced for ordinary purposes by a variety of scripts based on Italian models, but it is obvious that towards the end the scribes who used it had little confidence that everyone concerned would be able to read it and added an interlinear transcription in Arabic script. One of the manuscripts of the *Atebetü'l-Hakayik* is accompanied by such a transcription and it is perhaps significant that the two texts do not always agree.

Most of the surviving specimens of this script are literary texts, but one or two official documents in it have survived. These include the " yarlık " of Témür Kutluğ, dated A.D. 1398 and accompanied by an interlinear transcription in Arabic script, and the letter from Toxtamış to the King of Poland, dated A.D. 1397, of which facsimiles will be found at the end of Radloff's facsimile of the *Kutadğu: Bilig* cited above.

Perhaps the latest document in this script, no doubt written by a so-called *baxşı* employed at court for this purpose, and accompanied by an interlinear transcription in Arabic script, is Mehmed II the Conqueror's general proclamation of his victory over Uzun Ḥasan in A.D. 1473, published with a facsimile by R. R. Arat in *Türkiyat Mecmuası VI*, İstanbul, 1939, pp. 285 and foll.

The Buddhist religious M.O.A. continued to exist in the east, and one or two Mongolian Buddhist texts printed in China, no doubt for customers in the steppes, even as late as the last quarter of the fifteenth century, have recently come to light, but after the collapse of the (Mongolian) Yüan dynasty in A.D. 1368, the withdrawal of the Mongols to the steppes, and the development of almost complete anarchy in that area, Mongolian Buddhism, which was that of the old " Red Hat " sect, must have got into very

low water, and most of the Mongols seem to have reverted to shamanism, if indeed they had ever really abandoned it.

Towards the end of the sixteenth century there was a dramatic change in the situation. Early in the fifteenth century Tsong-kha-pa, the founder of the reformed "Yellow Hat" sect, had launched a revivalist movement in Tibet which gradually gathered strength in that country and by about A.D. 1566 spread to the Ordos Mongols in the great bend of the Yellow River. Translation of Buddhist texts into Mongolian, this time from Tibetan originals, soon started again, and this activity reached its peak when Ligdan Xan, who ruled from A.D. 1604 to 1634, gave orders for the whole of the Tibetan Kanjur and Tanjur to be translated into Mongolian. Prof. Walter Heissig has told me that, while most of this work was original, there is some evidence that when a Mongolian Buddhist text was already in existence, even if it was an old "Red Hat" one, it was used as the basis of the new edition.

But if some of the old Buddhist texts had survived, it is clear that the old orthography, and indeed probably the old language in which they had been written, had been almost completely forgotten, and what was in effect a new orthography was devised and a new and largely artificial language, which can be called "Classical, or Written, Mongolian," was based upon it. The essence of this new language was that while the old spellings were retained, at any rate to a large extent, many words, particularly those containing *gimel-cheth* and *caph*, were pronounced in a way in which they were not pronounced in the contemporary colloquial language and had not been pronounced at any earlier date. Thus, for example, "mountain" which had in the thirteenth century been pronounced *a'ula*, as we can see from the hP'ags-pa scription, and had been quite correctly spelt *aleph*, *gimel-cheth* (representing the inter-vocalic hiatus), *vau*, hooked *resh*, *aleph* in the old M.O.A., was still spelt in the same way but pronounced *ağula*.

The new M.O.A., with the phonetic values ascribed to its letters, can be found in any standard authority, for

example N. Poppe, *Grammar of Written Mongolian*, Wiesbaden, 1954.

This reconstituted alphabet is almost completely free from polyphonic letters, the only survivors being *aleph*, *vau* and *jod*, which were used as vowel letters in the traditional manner, *caph* used for both *g* and post-palatal *k*, and *lamed* and *tau* used indiscriminately for both *d* and *t*.

The principal points on which this reconstituted alphabet differed from the old M.O.A. were the following:—

(1) the old order of the alphabet had been forgotten and an almost completely new one was devised.

(2) the fact that the two superscribed dots had been used merely to identify *gimel-cheth* had been forgotten and also the various sounds represented by that letter; the letter, when undotted, was pronounced velar *k*, and when dotted *ğ*, that is in exactly the opposite way to the old Uyğur usage.

(3) the fact that there had been ways of representing *h* in the M.O.A., and indeed the very existence of *h* as a Mongolian sound, had been forgotten, and a completely new letter was invented to represent *h* in loan words.

(4) *zain* had disappeared apart from a vestigial survival as final *-s* " in old books."

(5) while medial *tzaddi* had remained recognizable, initial *tzaddi* had become indistinguishable from *jod*. A new outline was invented to represent initial and medial *ç*, the old form being retained to represent *c*. In recent times a new outline has been invented for initial *y-* to distinguish it from *c-*.

(6) *Beth*, which had been retained in the old religious M.O.A. for use in loan words like *vcir*, Sanskrit *vajra*, was still retained for use in such words.

(7) One or two new letters were invented by altering the outlines of existing letters to represent non-Mongolian sounds like *p* and *f* in loan words.

pai-tzü's, small metal badges of office which circulated throughout the Empire, and one or two of these have been found as far afield as southern Siberia. All the rest of the material comes from China. The total quantity is inconsiderable and some of it fragmentary. Almost the whole of it is assembled in N. Poppe, *The Mongolian monuments in hP'ags-pa script*, Wiesbaden, 1957, which also contains the text and a translation of a Chinese account of the alphabet which I discussed in the paper cited above.

The sounds which could be represented by this alphabet (no letter having more than one sound) can be tabulated as follows:—

VOWELS. The alphabet made provision for writing both short and long vowels, but the only long vowels which seem actually to occur in the texts are *a:* and *e:*, and it is not certain whether these are original long vowels or crases of two adjacent vowels (VV); it would therefore be better to leave open the question of long vowels. So far as the Mongolian texts are concerned (the position is much more complicated in the Chinese texts), the vowels represented in this alphabet were:—

Back *a*, *o*, *u*;
Front *e*, *é*, *ö*, *ü*;
Neutral *i*.

These texts are the only ones which provide clear evidence of the existence in early Mongolian of a closed *é*, which, except in certain diphthongs mentioned below, is found only in the first syllable of a word. The neutral vowel *i* occurs both in words containing back vowels and in words containing front vowels. There were special devices for representing the diphthongs ending in *-i* (or *-y*?); in *-ai*/*-ei* (or *-ay*/*-ey*) the final sound was represented by *-ü*, in those containing other vowels by *é*.

CONSONANTS. The table on opposite page shows the full range of sounds which could be represented by the letters of this alphabet, but those marked with an asterisk *

occur only in the Chinese texts or in Chinese and other loan words in the Mongolian texts.

	Plosive		Fricative		Nasal	Affricate		Sibilant		Semi-vowel
	v.	u.	v.	u.		v.	u.	v.	u.	
Labial	*b*	*p*,* *pb**	*v**	*f**	*m*	—	—	—	—	*w**
Dental	*d*	*t*,* *th*	—	—	*n*	*dz**	*ts*,* *tsh**	*z*	*s*	—
Denti-palatal	—	—	—	—	—	*ʒ*	*ʃ*,* *ʃh*	—	—	—
Palatal	—	—	—	—	*ñ**	—	—	*j**	*ʃ*	*y*
Post-palatal	*g*	*k*, *kh*	—	—	} *ŋ*	—	—	—	—	—
Velar	—	—	*ǧ*	*x**	} —	—	—	—	—	—

Liquids:—*l*, *r*; aspirate *h*; intervocalic hiatus '.

Post-palatal *k* hardly ever occurs in the Mongolian texts; it appears that the users of the alphabet heard a slight aspirate after all Mongolian unvoiced plosives and affricates. I discussed the question of velars at length in the paper cited above. Only one, which must represent a voiced fricative *ǧ*, occurs in Mongolian words both in this alphabet and in the Chinese transcriptions, but the Arabic transcriptions paint a different picture, so this must have been a peculiarity of the dialect spoken in China. The simple aspirate *h* is a fairly common initial sound, but in the intervocalic position it occurs only in *ihe-* " to protect " and *ihe:n* " protection " in the texts published by Poppe and apparently in *taha-* " to entrust " in a text not published by him. Voiced *z* occurs only in the word *zara* " a month," but this too must be a local idiosyncrasy, since there is good reason to suppose that this sound did not exist in Mongolian. Palatal *ʃ* occurs only in loan words, and before *-i-* in Mongolian words.

(b) THE TEXTS IN CHINESE TRANSCRIPTION

The only Mongolian texts in Chinese transcription which are important for the present purpose are the *Secret History*, the Chinese transcription of which, almost certainly from a text written in the M.O.A., was made by a commission of Chinese scholars between A.D. 1380 and 1390, and the Mongolian-Chinese vocabulary and handbook, the *Hua-I i-yü*, prepared by the same commission and published

in A.D. 1389. The first has been edited in a romanized text and translated several times, notably by Prof. E. Haenisch, and that scholar's index to it, *Wörterbuch zu Manghol un Niuca Tobcaan*, Leipzig, 1939, is the most convenient collection of the words which it contains. I shall have a good deal more to say about the second in the next chapter. Chinese script, even with the special devices which were employed in these texts to make the transcription more precise, is a very imperfect medium for representing a foreign language, and these transcriptions, invaluable though they are from other points of view and, so far as phonetics are concerned, for the light which they throw on such points as the initial *h-* and the various sounds represented by *gimel- cheth*, are not a reliable instrument for deciding on the finer points of phonetics. But it seems pretty clear, as I pointed out in the paper cited above, that the Mongolian dialect transcribed was identical with that presented by the hP'ags-pa texts.

(c) THE TEXTS IN ARABIC TRANSCRIPTION

Apart from isolated words and names in the works of Persian historians and the like, and in the *Sanglax* which is of little value for present purposes since it was written as late as the eighteenth century, there are three important early collections of Mongolian words in Arabic script:—

(i) The earliest is the Mongolian-Arabic glossary in the Leyden manuscript No. 517, Warner, which was written in A.D. 1245, probably in Persia or Transoxiana.* The wording of the colophon suggests that it was compiled by the scribe from oral information, and it therefore presumably represents the dialect spoken in that area in the middle of the thirteenth century. It formed the subject of a comprehensive study in N. Poppe, *Das mongolische Sprachmaterial einer Leidener Handschrift*, Izvestiya Akademii Nauk S.S.S.R., 1927.

* This manuscript also contains a Turkish-Arabic glossary published with an introduction and index by M. I. Houtsma in *Ein Türkisch-Arabisches Glossar*, Leiden, 1894.

(ii) Next comes the Mongolian-Arabic glossary in Camālu'l-dīn ibnu'l-Muhannā's *Ḥilyatu'l-insān wa ḥalbatu'l-lisān*. This work, which also contains a Turkish-Arabic and a Persian-Arabic glossary, cannot be precisely dated, but seems to have been written, probably in Iraq, early in the fourteenth century. The full text was published by Kilisli Mu'allim Rif'at in İstanbul in 1921, and an index of the words which it contains was included as an appendix to Poppe's work mentioned in (iii) below.

(iii) Some manuscripts of al-Zamaxşarī's *Muqaddimatu'l-Adab* contain voluminous glosses in Mongolian. The manuscripts concerned cannot be precisely dated but seem to belong to the first half of the fourteenth century. These glosses formed the subject of a comprehensive study in Poppe's *Mongolskiy Slovar' Mukaddimat al-adab*, Moscow-Leningrad, 1938 (cited as *Poppe* 1938).

The merits and demerits of Arabic script as a medium for representing foreign languages were discussed in Chapter V (1) (vi), and nothing more need be said on the subject here except that the demerits increase sharply when it is used, as it was in all these cases, by someone who is very imperfectly acquainted with the language which he is transcribing. The script was a very unsatisfactory one for representing vowel sounds, and little evidence of value on that subject can be extracted from these texts except that they contain a number of long vowels of all kinds (and not only *a:* and *e:* as in the ḫP'ags-pa texts) which seem to be mainly if not entirely secondary sounds due to crasis. The subject is discussed at some length in the Introduction to *Poppe*, 1938.

That Introduction also contains some valuable observations on the consonants used in these texts.

The distinction between the voiced fricative *ǧ* and the unvoiced plosive *k* is quite clear, and the use of one sound or the other in particular words corresponds reasonably well to the usage in the modern Mongolian languages, although there are some divergences. The unvoiced

fricative *x*, however, occurs only rarely and is clearly no more than a secondary form of *k*.

They can be tabulated as follows:—

	Plosive		Fricative		Nasal	Affricate		Sibilant		Semi-vowel
	v.	u.	v.	u.		v.	u.	v.	u.	
Labial	*b*	—	—	—	*m*	—	—	—	—	—
Dental	*d*	*t*	—	—	*n*	—	—	—	*s*	—
Denti-palatal	—	—	—	—	—	*c*	*ç*	—	—	—
Palatal	—	—	—	—	—	—	—	—	*ʃ*	*y*
Post-palatal	*g*	*k*	—	—	} *ŋ*	—	—	—	—	—
Velar	—	*k*	*ǧ*	(*x*)		—	—	—	—	—

Liquids:—*l*, *r*; aspirate *h*; intervocalic hiatus '.

In the medial and final positions there is a good deal of confusion between *ǧ* and *k*, but this seems to be no more than a local phenomenon, since there is exactly the same confusion between the two sounds in the Turkish languages, Çağatay and its predecessors, spoken in the same area. Poppe in his Introduction points out that in these texts velar as well as post-palatal sounds occur before *-i-*, although in "Classical Mongolian" only post-palatal sounds are permitted in this position. This seems to prove that in Mongolian *-i-* was always a neutral sound with front rather than back affinities. If a back *ı* had existed in early Mongolian as well as a neutral/front *i*, it is difficult to see why this consonantal sound change should have taken place.

Initial *h-* is common in these texts, but Poppe does not mention it as an intervocalic sound and I have not found any examples of it.

Both *c* and *ç* are represented by *cim* and the transcription of this letter by one or the other is a matter of arbitrary choice based on the pronunciation in other Mongolian languages.

ʃ very rarely appears in these texts except before *-i-* and it seems clear that words containing *ʃ* before other vowels are either loan words or secondary forms in which *ʃ* has been substituted for *ç*.

(d) THE EVIDENCE PROVIDED BY ARMENIAN AND GEORGIAN AUTHORITIES

A good many Mongolian proper names occur in the works of Armenian historians relating to the Mongolian conquest and subsequent events. As long ago as A.D. 1858 E. Dulaurier made a collection of these names and published it and a short Mongolian-Armenian glossary included in the History of Kirakos of Ganja (which covers events down to A.D. 1269 and must have been written soon after that date) in a series of articles entitled *Les Mongols d'après les historiens arméniens*, Journal Asiatique V, xi. Nearly a hundred years later a further list of Mongolian names and terms by F. W. Cleaves was published in R. P. Blake and R. N. Frye's *History of the Nation of the Archers* (*the Mongols*), Harvard University Press, 1954. It is probable that more material of the same kind still awaits collection.

An anonymous Georgian historian of the fourteenth century mentioned a few Mongolian words and expressions which were collected and studied by Prince I. A. Dzhavakov in an article published in the Izvestiya Akademii Nauk, Petrograd, 1917.

The methods of transcription used by these authorities is a fairly rough and ready one, and the material is inconsiderable, but seems to represent a dialect identical with, or very close to, that of the material in Arabic transcription.

(2) THE STRUCTURE OF THE WORD IN PRE-THIRTEENTH CENTURY MONGOLIAN

Considering how scanty the hP'ags-pa material is, and how unsatisfactory the early transcriptions in other scripts are, it is very difficult to analyze the structure of the Mongolian word in the way in which the structure of the Turkish word was analyzed in Chapter VI, and what I have to say in this section is intended rather to provoke thought among the Mongolists than to present a complete picture.

There is the same clear distinction in Mongolian as there is in Turkish between basic words and words elongated by the attachment of suffixes, and between verbs and

words which are not verbs, the latter, as in Chapter VI, hereafter called "nouns." There is also a great deal of evidence in Mongolian, probably even more than in Turkish, of an earlier stage of the language in which basic words still existed which are known to us only in an elongated form; for example there are many words like *tosun* "oil" carrying the nominal suffix *-sun/-sün* which have not survived without suffixes, and it is only from words like *de'edü* "sublime" and *de'ere* "above" that we can infer the existence of an earlier word *de'e*.

On the other hand, there do seem to be fundamental differences between the structure of Mongolian and Turkish:—

(1) There is some proclivity in Turkish for closed syllables; words ending in one or even two consonants (C or CC) are common, and also two consecutive consonants (CC) within basic words, which is another aspect of the same phenomenon. In Mongolian final CC is unknown, and words ending in consonants are less common than those ending with vowels. The typical Mongolian word is built up of a series of open syllables (V or CV), sometimes even as many as four, for example *bökö'üne* "gnat" in the *Hua-I i-yü*.

(2) Two consecutive vowels separated by an intervocalic hiatus (VV), which are unknown in early Turkish, are common in thirteenth century Mongolian. The situation has been greatly complicated by the fact that in the old M.O.A. VV was spelt V*gimel-cheth*(or *caph*)V, and in "Classical Mongolian" these combinations were invariably pronounced Vǧ(or *g*)V, but the evidence from the early transcriptions enumerated above shows that this is an error. It has been suggested that in this respect "Classical Mongolian" is more archaic than the thirteenth and fourteenth century dialects and preserves, in some mysterious way, sounds which had been lost in those dialects. Considering what we know of the circumstances in which the "Classical" alphabet came into existence, as described

in Chapter IX, this theory is really unrealistic and untenable. Moreover, it fails to explain why in some words, for example *çağa:n* " white " and *gege:n* " bright " V*ğ*(or *g*)V in " Classical Mongolian " goes back not to VV but to V*ğ*(or *g*)V in the earlier dialects. If there really was no intervocalic hiatus in pre-thirteenth century Mongolian, but only a *-ğ-/-g-*, which has survived in " Classical Mongolian " but was lost in the earlier dialects, it is difficult to see why it was preserved in some words in those earlier dialects and not invariably lost. Nevertheless there are some puzzling facts which have still to be explained, even if we quite rightly disregard the pronunciation of " Classical Mongolian " as having no value as evidence of the position prior to the thirteenth century. I pointed out in Chapter VIII that the balance of probability is that the voiced velar **ğ** in Turkish was a fricative, at any rate in the intervocalic position, although there is a possibility that in some positions, for example in a CC, it was a plosive. It is surprising that the thirteenth century Mongolian transcriptions of such Turkish loan words as **bağatur** and **kağan** are *ba'atur* and *ğa:n* (*ka:n*). (The latter is obviously a secondary form of *ka'an* and so supplies indirect evidence that at any rate in some cases Vv in thirteenth century Mongolian is a crasis of VV.) Nevertheless the fact remains that in these words the intervocalic hiatus in Mongolian corresponds to a fricative **-ğ-** in Turkish, and the possibility cannot be excluded that intervocalic fricative *-ğ-* was such a difficult sound for the Mongols in China in the thirteenth century to pronounce that they replaced it by an intervocalic hiatus, which is much the same thing as has happened in Turkey in recent years. If so, it is possible that there were in pre-thirteenth century Mongolian not only unvoiced velar and post-palatal *k* but also two voiced velar/post-palatal sounds:—

(a) A fricative *ğ/g* occurring only in the intervocalic position, which became a hiatus in thirteenth century Mongolian, thus producing two consecutive vowels, VV, which survive in the modern languages sometimes as VV,

but more usually by crasis as Vv, and (b) a plosive *ğ/g* occurring both as one component of a CC and also in the intervocalic position V*ğ/g*V, which still survives unchanged in the modern languages, except perhaps to the extent of becoming a fricative instead of a plosive in the intervocalic position.

(3) Whereas in Turkish most basic words are dissyllables, rather fewer are monosyllables, and there are very few basic trisyllabic nouns and no basic trisyllabic verbs, the position in Mongolian is that there are very few monosyllables, basic dissyllabic words are the commonest, basic trisyllabic nouns are quite common, and basic trisyllabic verbs not really uncommon.

(4) Whereas in early Turkish there are no unstable consonants at the ends of words (although in the mediaeval period final **-ğ** and **-g** became unstable and ultimately disappeared in many languages) in Mongolian from the earliest period known to us several consonants, and especially *-n*, were unstable in the final position. This is no doubt another aspect of the Mongolian dislike for closed syllables referred to in (1) above. There seem to be at least two stages of instability. The plural of the adjectival suffix *-tay/-tey*, in which the *-y* is unstable, is *-tan/-ten*, in which the *-n* could hardly have been unstable if the singular and plural were to be distinguished. Again in *dörben* " four " the final *-n* seems to be stable, but is elided when the ordinal and other suffixes are attached. On the other hand, " disease " seems to be spelt *ebeci* as often as it is spelt *ebeçin*, so that it is quite uncertain whether this word should be classified as VCVCV or VCVCVC. It is possible that in the latter case the *-n* is not part of the original word but merely a sort of euphonic affix. This is to some extent supported by the fact that the Turkish word **élçi**: " envoy," which certainly had no final **-n**, appears as a loan word in Mongolian both as *élçi* and as *élçin*. Poppe explains the second as a Mongolian plural in *-n* of *élçi*, but it is difficult in some cases to see that it has any plural meaning.

Subject to these remarks, I suggest that the forms of the Mongolian basic word can be classified as follows. For the reasons given above the list may not be complete, and some of the longer forms may in fact be those of words which are not basic but carry suffixes which I have failed to recognize. I have not included in this list any forms which I have found only in the case of Turkish or other loan words.

MONOSYLLABLES

Nouns. I have found only one word of the form VC *ür* " dawn " (*Poppe*, 1938). There are several words of the form CV, mostly particles like *ba* " and," or pronouns like *bi* " I." There are rather more words of the form CVC; I shall discuss below the final sounds in such words.

Verbs. There seem to be rather more forms. For V- there seems to be only *a-* " to be "; for VC- I have found only *ab-* " to take " and *ög-* " to give," and for CV- only *bö-* " to be " and *ki-* " to make," but there may be a few more. There seems to be, at any rate one verb *ge:-/ke:-* " to say " of the form CVv-. The only form which is rather commoner is CVC-; most of these verbs seem to end in *-r-*.

DISSYLLABLES

Nouns. I have noted nouns of the following forms:—

—	CVV
VVC	CVVC
VCV	CVCV
VCCV	CVCCV
VCVC	CVCVC
—	CVCVvC
VCCVC	CVCCVC

I have not found many words containing VV in any position and still fewer containing Vv. By far the commonest form is CVCVC, but as in some of them the final C is an *-n*, probably unstable, these latter words should perhaps be classified as CVCV. Similarly some of the words of the form CVVC should perhaps be classified as CVV.

Verbs. I have noticed verbs of the following forms:—

VV-	CVV-
VVC-	—
VCV-	CVCV-
VCCV-	CVCCV-
VCVC-	CVCVC-

Of these by far the commonest are VCV- and CVCV-; there do not seem to be many verbs ending in -C-; the only verb of the form VV- which I have found is *u'u-* "to drink."

TRISYLLABLES

It is much more difficult to classify the trisyllabic words, partly because the possible forms are much more numerous, and partly because words of this length are more likely to contain unrecognized suffixes.

Nouns. I have noticed nouns of the following forms:—

VCVV	(CVCVV)
VVCV	—
VCVVC	CVCVVC
VVCVC	CVVCVC
VCVCV	CVCVCV
—	CVVCCV
VCCVCV	CVCCVCV
VCVCVC	CVCVCVC
VCVCCVC	CVCVCCVC
VCCVVC	CVCCVVC

Of these the commonest form is perhaps CVCVVC, but a good many words of this and other forms end with an unstable *-n* and should perhaps be classified without the final C.

Verbs. There seem to be fewer forms, and I have not noticed any basic trisyllabic verbs with a final -C-. The forms which I have noticed are:—

VCVCV-	CVCVCV-
—	CVCVV-
—	CVVCV-

(3) THE SUFFIXES IN MONGOLIAN

The Mongolian suffixes fall into the same classes as the Turkish suffixes. I explained these classes in Chapter VII, and also mentioned those classes—the possessive, declensional and conjugational suffixes and the post-positions —to which I do not propose to refer.

Poppe's *Grammar of Written Mongolian*, Wiesbaden, 1952, contains several lists of suffixes of various kinds, all of them listed again in alphabetical order in the Index. The lists, of course, relate to "Classical Mongolian." Most of these suffixes obviously go back to beyond the thirteenth century. Some of them, even of those as old as the thirteenth century, are certainly, and some possibly, borrowed from Turkish, and used chiefly with Turkish loan words. Others may, on the analogy of Turkish, have come into existence later than the thirteenth century. On the other hand, there is at least one suffix which was already unproductive by the thirteenth century, the deverbal nominal suffix *-s*, which can be inferred from a comparison between *heçüs* "end" and *heçül-* "to come to an end" in the hP'ags-pa texts, and there may be others awaiting discovery. As a first step, however, it will, I think, be useful to arrange the suffixes in Poppe's list in the same classes, and in the same alphabetical order, as the Turkish suffixes in Chapter VIII. As the functions of these suffixes are fully explained in Poppe *op. cit.*, it is unnecessary to repeat the information here, except when special points arise, but I have marked those suffixes which form adverbs "Adv.," and mentioned the functions of the deverbal verbal suffixes, which are interesting because they differ quite substantially from those in the equivalent list of Turkish suffixes. I have also marked:—

(1) those which are certainly borrowed from Turkish with "L" (for "loan").

(2) those which are possibly borrowed from Turkish with "L?".

(3) those which I have found in the early texts with "13th" (for the hP'ags-pa inscriptions and the Secret History) or "14th" (for the HIiy).

Where a suffix in "Classical Mongolian" contains an intervocalic *-ğ-/-g-* I have left it unchanged except when I have found it in the early texts as an intervocalic hiatus, id which case I have substituted *-'-*.

(1) Nominal suffixes

(a) Denominal

-a/-e (Adv.); *-'a/-'e* (Adv., 13th); *-da/-de* (Adv., 13th); *-ta/-te* (Adv.); *-na/-ne* (Adv., 13th); *-ğana/-gene* (13th); *-ra/-re* (Adv., 13th).

-çi (L. 13th)*; *-bçi* (14th); *-ki* (L.); *-bki; -daki/-deki/-taki/-teki* (L); *-ğali/-geli; -şi* (Adv.); *-ğşi/-gşi* (Adv., 13th).

-çu/-çü (13th); *-du/-dü/-tu-/-tü* (13th); *-ru/-rü* (Adv., L?, 13th).

-mad/-med (13th).

-rkağ/-rkeg (possibly deverbal suffix *-ğ/g* attached to Turkish verbal suffix *-rğa:-/rge:-*, see Chapter VII (2) (a)); *-mağ* (Diminutive, not connected with Turkish **-ma:k/-me:k**); *-sağ/-seg* (L.); *-liğ/-lig* (L.); *-msuğ/-msüg.*

-tan/-ten (13th); *-kan/-ken* (14th); *-cin* (13th); *-lcin; -ğçin/-gçin* (13th); *-lun/-lün* (13th); *-sun/-sün* (13th).

-laŋ/-leŋ.

-çar/-çer (forms adjectives and adverbs); *-nçar/-nçer* (? 13th in the proper name *Bodonçar*); *-ğar/-ger* (Adv., 13th); *-sar/-ser* ("first layer" loan suffix from **-sız/-siz**); *-bir; -bur/-bür; -btur/-btür; -ğur/-gür* (perhaps for *-'ur/-'ür*).

-tay/-tey (13th); *-ğtay/-gtey; -kay/-key; -ğuy/-güy* (dubious, perhaps inferred from a Turkish loan word).

(b) Suffixes attached to numerals†

-ta/-te forms numeral adverbs:—"once" *etc.* (13th).

-'ula(n)/-'üle(n) collective (13th; borrowed in some Turkish languages).

* Poppe, para. 356, also mentions a Nomen Actoris in *-ğçi/-gçi*, but this seems to be merely a combination of deverbal *-ğ/-g* and *-çi*; in paras. 147 and 269 he also mentions a suffix *-ğaçi/-geçi*; this might be a "first layer" borrowing of the Turkish deverbal suffix **-ğu:çı:/-gü:çi:**, see Chapter VII (1) (c), or a longer form of *-ğçi/-gçi* with euphonic vowel inserted. None of these suffixes are connected with *-ğçin/-gçin*, the feminine suffix of adjectives, Poppe, para. 120.

† It will be noticed that Mongolian has a numeral adverb, which is completely lacking in Turkish.

-ğad/-ged (perhaps for *-'ad/-'ed*) distributive.
-kan/-ken diminutive:—" only one " *etc.*
-du'ar/-dü'er ordinal (13th).

(c) Deverbal

-ca/-ce (13th); *-ça/-çe; -lta/-lte* (13th); *-mta/-mte; -lğa/-lge; -ma/-me.*
-mci; (*-ğçi/-gçi* and *-ğaçi/-geçi*, see note on *-çi* in (1) (a)); *-li* (13th); *-ğuli/-güli* (perhaps for *-'uli/-'üli); -mi; -ri* (13th); *-buri/-büri* (13th); *-ğuri/-güri* (perhaps for *-'uri/-'üri); -şi* 13th).
-'u/-'ü (13th).
-ğ/-g (L., 13th); *-mağ/-meg* (L.); *-mşiğ/-mşig.*
-l; -dal/-del (13th); *-mal/-mel* (14th); *-'ul/-'ül* (13th).
-m (L.).
-n; -'un/-'ün (13th); *-dun/-dün; -sun/-sün* (13th); *-'asun/-'esün* (13th); *-dasun/-desün.*
-ŋ; -laŋ/-leŋ (13th); *-kulaŋ/-küleŋ* (*-laŋ/-leŋ* attached to infinitive in *-ku/-kü;* 13th).
-r (13th); *-'ar/-'er* (13th); *-mar/-mer; -msar/-mser; -'ur/-'ür* (13th).
-s (13th; see introductory remarks).
-ğay/-gey (13th); *-mağay/-megey/-mkay/-mkey; -ŋğuy/-ŋgüy.*

(2) Verbal suffixes

(a) Denominal

-a-/-e- (not in Poppe, but attested by *delgere-* < *delger* in hP'ags-pa texts); *-da-/-de-/-ta-/-te-* (13th, possibly a secondary form of *-la-/-le-*, if so, L.); *-rka-/-rke-* (? L.); *-la-/-le-* (13th, L.); *-çila-/-çile-; -gina-/-gine-* and *-çigina-/-çigine-* (both attached only to onomatopoeics); *-ra-/-re; -kira-/-kire-* (attached only to onomatopoeics); *-şiya-/-şiye-* (13th).
-ci-; -çi- (attached only to adverbs); *-gi-* (attached only to onomatopoeics); *-şi-* (13th).
-tu-/-tü-.
-V*d-* (13th).
-l- (attached only to adverbs).

(b) Deverbal

-'a-/-'e- causative (13th).
-lca-/-lce- Durative.
-balca-/-belce-/-ğalca-/-gelce- Durative.
-lça-/-lçe Cooperative (13th).
-da-/-de-/-ta-/-te- Passive (13th).
-ğda-/-gde- Passive (13th).
-ğa-/-ge-/-ka-/-ke- Causative (13th).
-çağa-/-çege- (perhaps *-ça'a-/-çe'e-*) Plural.
-lğa-/-lge- Causative (13th).
-ra-/-re- Reflexive (13th). [illegible]
-ldu-/-ldü- Reciprocal (13th). [illegible]
-l- Iterative, and (obsolete) in some cases Causative (13th).
'ul-/-'ül- Causative (13th).

(4) THE RECONSTRUCTION OF THE PHONETIC STRUCTURE OF PRE-THIRTEENTH CENTURY MONGOLIAN

The evidence which has been collected above gives a fairly clear picture of the actual sounds which existed in thirteenth century Mongolian, and enables us to infer some changes which must have taken place at some time before that date, but the exact phonetics of some Mongolian words still elude us, and it is not easy to be certain what consonantal sounds could be associated in CC clusters.

VOWELS. There is no reasonable doubt that the vowels which existed in pre-thirteenth century Mongolian were:—

Back *a, o, u.*
Front *e, é, ö, ü.*
Neutral *i.*

The hP'ags-pa texts make it clear that closed *é*, which could occur only in the first syllable except as a scription for *-i* (or *-y*) at the end of certain diphthongs, was commoner as an initial than open *e-*, at any rate in the dialect there represented, and also occurred after an initial consonant, but there is some possibility that this is a slight distortion of the earlier situation. Initial closed *é-* occurs in some loan words in which there was the same initial sound in

Turkish, but also in some words which had an open **e**, for example *érdem* < **erdem**, or even an *i*, for example *éce(n)* < idi:. As there is no other evidence on the subject we cannot say with confidence in Mongolian as we can in Turkish, in what words closed **é** occurred. I stated in section (1) (c) above the reasons for which I do not think that we can infer the existence of more than one *i* sound in pre-thirteenth century Mongolian. I have also given the reasons for which I think that it is not at present possible, and perhaps never will be, to decide whether there were originally long vowels in Mongolian.

Leaving the neutral *i* on one side, it is clear that in pre-thirteenth century Mongolian a word could not contain both back and front vowels; there is no evidence, as there is in Turkish, that certain suffixes originally had stable vowels of one sort or the other, irrespective of the quality of the vowels in the words to which they were attached; and there does not seem to be any evidence that there were rules determining what vowels in later syllables should follow particular vowels in the first syllable, except perhaps that if the vowels both in the first and later syllables were rounded, which was not necessarily the case, they were all either *o*, *u*, *ö* or *ü* and not a mixture.

CONSONANTS. All the consonantal sounds which existed in pre-thirteenth century Mongolian could occur in the medial position but some could not occur as initials and others as finals.

Medial. The only labial sounds were *b* and *m*. *P* and *v* in loan words became either *b* as in *arba* < **arpa:**, or *b* as in *cibiyin* < **cipkin** < **yipkin**, or *ğ* as in *ciğar* < ***cıpa:r** < **yıpa:r**, or an intervocalic hiatus as *şi'ür-* < **süpür-**. Exceptionally however in early Buddhist texts Turkish loan words containing *v* were taken over unaltered so far as their written form was concerned.

The only dental sounds were *d*, *t*, *n* and *s*. The fricative *d̲* in loan words became *d*. There is one peculiarity in the pronunciation of even thirteenth century Mongolian, that

di and *ti* did not exist, but were replaced by *ci* and *çi;* this change was invariably made in loan words.

Another peculiarity of even thirteenth century Mongolian was that *si* did not exist, and this combination, if it occurred in a loan word, became *ṣi*. There is however this difference between this case and the last, that while *d-*, *t-*, *c-* and *ç-* all occur before other vowels, *ṣ-* does not occur before any vowel except *i* save in loan words. The obvious conclusion is that *ṣ* did not originally exist in Mongolian and only evolved later as a result of palatalization by a following *-i*. This is confirmed by the fact that the Turkish word **uluş** " a country," a second layer loan word, became *ulus* in Mongolian.

As regards the velar sounds it seems clear that there were in pre-thirteenth century Mongolian an unvoiced plosive *k*, a voiced fricative *ǧ* and possibly a voiced plosive *ǧ*, but not an unvoiced fricative *x*. The position of the post-palatals is less clear, but there does not seem to be any evidence of the existence of any sounds except the plosives *g* and *k*.

The nasal sound *ŋ* occurred in association both with velars and with post-palatals. It could not exist as a pure intervocalic sound; intervocalic **-ŋ-** in Turkish loan words became *-ŋǧ-* or *-ŋg-*.

There is ample evidence of the existence of an intervocalic hiatus in thirteenth century Mongolian, but we cannot at present say whether this had always been a feature of the language or was a secondary phenomenon. The aspirate *h* occurs so rarely in the medial position, even in thirteenth century Mongolian, that it seems doubtful whether it was not a secondary sound in the few words in which it occurred. The question of initial *h-* is discussed below.

Thus the consonantal sounds of pre-thirteenth century Mongolian can be tabulated as shown on opposite page.

The question which of these sounds could occur as components in a consonantal cluster (CC) is complicated by the scantiness of the hP'ags-pa texts and the unreliability

	Plosive		Fricative	Nasal	Affricate		Sibilant	Semi-vowel
	v.	u.	v.		v.	u.	u.	
Labial	*b*	—	—	*m*	—	—	—	—
Dental	*d*	*t*	—	*n*	—	—	*s*	—
Denti-palatal	—	—	—	—	*c*	*ç*	—	—
Palatal	—	—	—	—	—	—	—	*y*
Post-palatal	*g*	*k*	—	} *ŋ*	—	—	—	—
Velar	ʔ *ğ*	*k*	*ğ*		—	—	—	—

Liquids:—*l*, *r*; aspirate *h*; intervocalic hiatus '.

of the rest. It is in any event clear that a syllable could not end in CC, so that neither final CC nor medial CCC could occur in a native word; when CCC occurred in a loan word an euphonic vowel was inserted, for example **arsla:n** became *arsalan*. In the following table, which shows all the combinations which I have found in thirteenth and fourteenth century texts, I have not included CCs of which the second C is the first sound of a suffix. I have also excluded *ş* from the table, although this sound does occur in the various combinations C*şi*.

	b	*c*	*ç*	*d*	*t*	*ğ*	*k*	*g*	*k*	*l*	*m*	*s*
b	—	—	*bç*	*bd*	*bt*	*bğ*	*bk*	*bg*	*bk*	—	—	*bs*
d	—	—	—	—	—	—	*dk*	—	—	—	—	—
t	—	—	—	—	—	—	*tk*	—	—	—	—	—
ğ	—	—	*ğç*	*ğd*	*ğt*	—	—	—	—	—	—	*ğs*
g	—	—	*gç*	—	—	—	—	—	—	—	—	—
l	*lb*	*lc*	*lç*	*ld*	*lt*	*lğ*	*lk*	*lg*	—	—	—	—
m	—	*mc*	—	*md*	—	*mğ*	*mk*	—	*mk*	—	—	*ms*
n	—	*nc*	—	*nd*	—	—	—	—	—	—	—	—
ŋ	—	—	—	—	—	*ŋğ*	*ŋk*	*ŋg*	*ŋk*	—	—	*ŋs*
r	*rb*	—	*rç*	*rd*	—	*rğ*	*rk*	*rg*	*rk*	—	*rm*	—
s	—	—	—	—	—	*sğ*	—	*sg*	—	—	—	—
y	—	—	—	—	—	—	—	—	—	*yl*	*ym*	—

The list is perhaps not quite complete and one or two entries are perhaps doubtful. *T* is the only unvoiced first component plosive of a CC, and the only combination in which it occurs, *tk*, may be an error for, or secondary form of, *dk*. Owing to the obscurities of the scripts concerned, it is doubtful whether *yl* and *ym* are true CCs and not faulty scriptions for *yil* and *yim*. In any event the list is markedly shorter than the corresponding Turkish list in Chapter VIII, containing about fifty items against seventy. What is

even more surprising is the difference between the two lists. In this list there are at most twelve, and perhaps only ten, possible first components of a CC as against nineteen in the Turkish list, but of those ten (or twelve) one (or two) do not occur in the latter. One of the commonest first components here is *b*, which occurs in the Turkish list only in **br,** a combination not found in this list. Conversely **-r-,** one of the commonest second components in the Turkish list, is completely lacking in this list, and that this is no accident is proved by the fact that when such a combination occurred in a loan word an euphonic vowel was inserted; for example **köprüg** became *ke'ürge*.

Initial. The only sounds occurring in a medial position which are not found as initials are *l*, *ŋ*, *r* and of course, by definition, the intervocalic hiatus. As a result of the great interest which has been taken in the subject during the past fifty years, there can be no doubt that *h-* was a fairly common initial in thirteenth century Mongolian and still survives, in one form or another, in several modern languages. Nevertheless there is reasonable doubt whether this was originally a Mongolian sound. The phenomenon of secondary aspiration is a fairly familiar one in a good many languages, including for example some of the Turkish languages, and notably the Türki (or "Neo-Uyğur") of Sinkiang. There is no doubt at all that the initial **h-** in these Turkish languages is a secondary phenomenon, and equally there is no doubt that in at any rate some Mongolian words this is true also. For example *hüker* "an ox," which incidentally appears as *üker* in *Poppe*, 1938, page 377, although there are many initial *h*-s in the same text, is undoubtedly a first layer loan word from **öküz,** which certainly never had an initial **h-**. Conversely "thus" is spelt *éyin* in the hP'ags-pa texts, no doubt correctly, but *heyin* in *Poppe*, 1938, page 183. The matter is obviously one which requires further examination, but my present view is that on balance the initial *h-* is more likely to be a secondary phenomenon than an original feature of the language.

Final. It is clear that in the thirteenth century the unvoiced plosives *t*, velar and post-palatal *k* and the affricate *ç* could not occur at the end of a word. The clearest evidence of this is the orthography of the hP'ags-pa texts; the Chinese transcriptions are useless on this point, and the Arabic scriptions, as pointed out above, are vitiated by a widespread tendency to devoice voiced finals in the area where they were written down. The point is also proved by the treatment of loan words. If the Sanskrit loan words *adhiṣṭita* and *Bodhisattva* appear in the hP'ags-pa texts as *adişdid* and *bodhisiwid* the most reasonable explanation is that the Mongols could not at that time pronounce a final *-t*. There is similar evidence in the treatment of Turkish loan words with final unvoiced sounds (see Chapter XI). Similarly *h* and, by definition, the intervocalic hiatus could not occur at the end of a word.

There is no theoretical reason why the remaining twelve sounds, *-b*, *-c*, *-d*, *-ğ*, *-g*, *-l*, *-m*, *-n-*, *-ŋ*, *-r*, *-s* and *-y* should not occur at the end of a word, but I have not found any native word, basic or elongated, ending in *-c*. Monosyllables are, of course, rare, but I have found monosyllabic basic nouns ending in all the remaining eleven sounds except *-d* and *-ŋ*, and monosyllabic basic verbs ending in *-b-*, *-d-*, *-g-*, *-r-* and *-y-*, but of these the only common ending is *-r-*. So far as longer words are concerned the position is complicated by the facts that all these eleven sounds except *b* and *y* occur as suffixes and except *b* and *s* at the end of suffixes. The position however appears to be that there are a good many dissyllabic basic nouns ending in *-n* (both stable and unstable), *-r* and *-y* and a few ending in *-b* and *-l*, but I have not found any ending with the other six sounds. I have found dissyllabic verbs ending in *-d-*, *-l-* and *-s-*: those that end in *-s-* must be basic, but those that end in *-d-*, like *e'üd-*, *usad-* and *deled-* or in *-l-* like *kağal-* may be either basic or suffixed forms of obsolete basic verbs.

There seem to be some trisyllabic nouns ending in *-l* and *-n* (both stable and unstable) which are certainly basic,

and a few ending in *-r* and *-y* which might be basic, but I have not found any trisyllabic basic verbs ending in consonantal sounds.

The whole subject of these final sounds is a complicated one, and really requires examination by some scholar with a much greater knowledge of Mongolian than I possess.

CHAPTER XI

THE RELATIONSHIP BETWEEN TURKISH AND MONGOLIAN

In Chapter II I explained how the various primitive human groups evolved languages of their own, and how when a group split up into sub-groups the primitive language of the group broke up first into dialects and then, as time went on, into separate languages, and I illustrated this process by explaining how the original unitary Turkish language first broke up into a " standard " and an " l/r " language, and how each of these in course of time split up again into separate languages, the descendants of which form the family of Turkish languages to-day. I pointed out that other families of languages—Indo-European, Semitic, Uralian and so on—had evolved in exactly the same way. All this is, of course, no more than an elementary statement of fact and has been said many times before, but the repetition of elementary facts, even if it is sometimes irritating, does at any rate promote clarity of thought.

If it were possible, as of course it is not, to trace the whole history of a family of languages, something like the genealogy of a human family could be constructed. At the top would be the original unitary language of the primaeval group, the next generation would be constituted by two or three languages which evolved directly from that primitive language. In the next generation similarly each of these first generation languages would have several descendants, and so on until we got to the languages of the family which are still spoken to-day. All the rest would be dead, but some would still be preserved in written documents; others would have disappeared entirely. Assuming, again contrary to the possibilities, that all these languages from the very earliest onwards had survived in written documents, it would be possible to work out in detail the history of the vocabularies of all of them. The

vocabulary of the original language would be found to be a very restricted one, amounting to no more than a few words for simple things and ideas and, at any rate in its earlier phases, possessing no kind of grammatical structure. The languages of the next generation would each preserve the whole, or nearly the whole, of that ancestral vocabulary, but would add new words to it, each in its own way, by one of the five methods enumerated in Chapter II. The same process could be traced from generation to generation. Some of the words used by earlier generations would be found to have fallen out of use, and of those that survived some would have undergone phonetic or semantic changes, but generally speaking it would be found that in each generation more of the vocabulary of the last generation had survived than had been lost.

If we reverse the process and examine the vocabularies of all the languages belonging to a particular family that are known to us, we are able to some extent to reconstruct the vocabularies of earlier members of the family, and even of languages which have been completely lost. What we find is that each modern language has some words peculiar to itself and acquired since it became a separate language by one of the five methods already referred to. It may also during this period of separate existence have acquired loan words from other languages in the family. It is important to identify these as recent loan words and not use them as evidence of a genetic connection between the languages concerned. For example within the past two or three hundred years English has borrowed "pyjamas" from Hindustani. This word has no evidential value; what does prove a distant genetic connection between the two languages is that Hindustani *py-* (*pai*) and English *foot* are both descended by different lines of descent from a very early Indo-European word meaning "foot" or "leg."

Below this top layer of words peculiar to individual members of the group we find a layer of words which are common to several languages that had a common ancestor in the last generation (for example in the Indo-European

family the Germanic, Romance or Slavonic languages) but are not used, except as loan words, in languages which do not belong to the particular cluster of languages concerned.

Below this again we find a third layer, a smaller number of words common to two or more of the clusters of languages referred to in the last paragraph and inherited by them all from a common ancestor in an earlier generation.

And finally we may be able to isolate the lowest layer of all, words now varying greatly in sound, and perhaps to some extent in meaning, which exist in nearly every language of the family and were inherited from the original primitive language. The words in this lowest layer are all the kind of words which would be required by a very primitive society, simple verbs and nouns of action or state, names of everyday objects, parts of the body, terms of relationship, numerals and the like. In the next layer are words for rather more advanced concepts, the need for which are felt in a more advanced society, and so on. If the words in each layer, except usually the lowest, are examined in order to determine which of them were inherited or produced by spontaneous generation or deliberate creation, and which of them were borrowed from some other language, and if it is possible to identify that language, it can safely be assumed that the borrowed words are words for things or ideas borrowed from the speakers of that language, and not previously known to the speakers of the borrowing language.

In examining the more modern languages of a group we have to take account of another phenomenon. When one people speaking its own language is subjugated by a people speaking a different language and has its rule imposed upon it, whether the dominant people is more or less advanced in civilization than the subject people, the subject people will have to take into its own language a good many of the words of the dominant people, not because it has no equivalent words of its own but because the dominant people expects its subjects to understand and use these words in its dealings with it. At the worst the

subject people may lose its own language completely; at the best, if it is more advanced than the dominant people, it may find that, although it has to accept a good many words for which it has no real use, the dominant people will accept a good many of its words for things and ideas for which it has no words of its own. The same kind of thing happens when a small group of people speaking a particular language lives for a long time isolated from its kinsmen and surrounded by people speaking a different language. A case in point is the Moghols of Afghanistan, who are the descendants of a small group of Mongols, probably an army or a garrison, which was left behind in Afghanistan in the thirteenth century, and has lived there ever since surrounded by Afghans. The vocabulary of this people, who are and presumably always were completely illiterate, has recently been analysed by P. Poucha in *Die Sprache der Mogholen in Afghanistan und die Theorie der Mischsprachen*, Central Asiatic Journal, VI, 1, 1961. The analysis shows that, while the grammar of the language, the pronouns and the verbs remain almost purely Mongolian all other sections of the vocabulary, including even the numerals, are saturated with Iranian loan words.

It is, therefore, possible by the intelligent use of vocabulary analysis and history to determine with reasonable certainty whether two languages having a certain number of words in common have them in common because they have been inherited by both languages from a common ancestor, which proves that the two languages belong to the same family of languages, or whether they have them in common because they have been borrowed by one language from the other, which proves that the two languages belong to different families.

If the common words, apart from recognizable recent loan words, are the kind of words which are likely to have been used in both languages from the most primitive period, then it can safely be assumed, even if such words form only relatively small parts of the whole vocabularies, that the two languages are descended from a common

ancestor. Indeed the proportionate bulk of such words in the total vocabularies may indicate how remote that common ancestor is. But if the basic vocabularies are entirely different, and the common words are the kind of words which would not have been required until the societies concerned became less primitive, or if the common words do not fall entirely within this class, but there is historical evidence that the speakers of the two languages have at some time been in very close contact, and especially if the speakers of one language have been dominated by speakers of the other, then it can safely be assumed that the languages in question belong to different families.

A comparison of modern languages which are generally accepted as belonging to the Turkish family of languages shows that in the case of those which have not been in contact with one another for many centuries the common words are of a kind likely to have been in use from the earliest period. This is true, for example, of Yakut and Chuvash, languages with very different phonetic structures spoken at opposite ends of the Turkish-speaking world by peoples whose ancestors must have lost contact with one another not later than the fifth century A.D. and probably a good deal earlier. In Chuvash, for example, " five " is **pilĕk,** " daughter " **xer** and " water " **şiv**; in Yakut the corresponding words are **bies, kı:s** and **u:**; at first sight the two sets of words seem to be completely different, but by studying the phonetic structure of the two languages and the languages from which they are descended it is easy to show that both sets of words are descended from the forms which these words must have had in the original unitary language, and which can be reconstructed as **bé:ş, kı:z** and **su:v.**

According to the Altaic theory the original unitary Turkish language was not an independent primaeval language but, together with the original unitary Mongolian and Tungus languages, a " first generation " descendant of an older unitary language provisionally called " Proto-Altaic." An admirable account of the origin and develop-

ment of the Altaic theory, together with an exhaustive list of the authorities in which it has been expounded, will be found in J. Benzing, *Einführung in das Studium der altaischen Philologie und der Turkologie*, Wiesbaden, 1953, pages 1 and foll. I have discussed it at length in three papers in the Central Asiatic Journal and in *Turk, Mongol, Tungus*. These papers may have some historical interest as showing how my ideas on the subject developed; they contain some valid arguments and some mistakes, but there is a good deal of repetition in them and they do not amount to a coherent and logically arranged examination of the theory. It would I think be useful to attack the whole subject again *de novo*.

It seems to me that the Altaic theory stands or falls on one single point. If the vocabularies of the earliest available representatives of the Turkish, Mongolian and Tungus families of languages are analysed and it is found that the basic vocabularies of the three, that is the kind of words which are likely to have been in continuous use from the most primitive period are all entirely different, and that the words which they have in common are the kind of words which less advanced peoples might be expected to borrow from more advanced peoples with whom they came in contact, then the Altaic theory cannot be valid.

So far as the Tungus languages are concerned, for the reasons stated in *Turk, Mongol, Tungus* I do not think that sufficient material has yet been assembled to make a detailed examination possible, but *prima facie* the basic Tungus vocabulary is quite different from that of Mongolian and Turkish and the common words are of the kind which might be expected to be loan words.

As regards Mongolian, however, much more comparable material is available, and I think that it can be demonstrated beyond any possibility of doubt that the basic vocabularies of Mongolian and Turkish are entirely different and that the words which thirteenth and fourteenth century Mongolian had in common with Turkish languages of the same or an earlier date were borrowed from the Turks by the Mongols. The compilers of the Chinese-

Mongolian vocabulary called the *Hua-I i-yü* "the Chinese and Barbarian Interpreter" (cited as *HIiy*), published in A.D. 1389, arranged their material in seventeen sections or chapters in the most convenient way possible for facilitating a comparison between the Mongolian and Turkish vocabularies. The *HIiy* is not itself very extensive, containing only 846 entries, apart from the specimen Mongolian texts in the second part from which further words can be extracted, and it is not a list of Mongolian words with Chinese translations but a list of Chinese words with the nearest Mongolian equivalents that the compilers could find, so that some of the equivalents are approximations rather than exact translations, while others show merely that a particular Chinese word had no exact Mongolian equivalent and had to be paraphrased. It is, however, fairly easy to add other words to the individual sections taken from other contemporary authorities, and so broaden the basis of comparison.

It is quite easy to demonstrate that many of the words in the vocabulary are loan words, mostly taken from Turkish, and it also seems to me quite often possible to estimate the date at which they entered the Mongolian language, and specifically that there are three layers of loan words borrowed at different periods from different Turkish languages.

The first or lowest layer, which I have suggested contains words borrowed before the eighth century A.D., and probably in the fifth or sixth century by the Kitan from the Tavğaç, includes words which retain, somewhat modified, an aspect more archaic than that of the corresponding Turkish words in any but a few Turkish languages known to us. It is obvious that when there are phonetic differences between a loan word in Mongolian and the equivalent word in "standard" Turkish the reason may be either that the form of the word in the language from which it was borrowed was different from its form in "standard" Turkish, or that the word, when borrowed, had to be altered to suit Mongolian habits of pronunciation,

or indeed both, and it is often not easy to distinguish between these two kinds of difference. It seems clear that the words in this layer were borrowed from a language which retained certain archaic traits that had disappeared from Türkü and Uyğur, but at the same time had undergone some phonetic changes which had not occurred in those languages. The phonetic differences between this language, which was I suggested Tavğaç, and, say, Türkü can be summarized as follows:—

(1) Initial **d̲-** survived in words in which it had become **y-** in Türkü;

(2) initial **ñ-** survived in words in which it had become **y-** in Türkü;

(3) initial **g-** survived both in this language and in some Oğuz languages in words in which it had become **k-** in Türkü;

(4) medial and final **ş**, which was retained in Türkü, had become **l**;

(5) medial and final **z**, which was retained in Türkü, had become **r**;

(6) **-ğ-/-g-** forming one component of a CC in Türkü words had been elided;

(7) intervocalic **-ğ-/-g-** found in Türkü words had become **-y-** (at any rate before **-ı-/-i-**).

(8) initial **b-** had become **m-** in all words containing nasals; initial **b-** was retained in this position in the earliest Türkü and Uyğur texts, in the Tuvan inscriptions (Old Kırğız) and in some languages of the Oğuz group; the change to **m-** which began in the later Türkü and Uyğur texts is now almost universal.

(9) various vowel changes, including medial **-ı-/-i-**, **-o-/-ö-** and **-u-/-ü-** > **-a-/-e-** had occurred; similar vowel

changes have been observed in Uyğur-A but not in any other language.

It will be observed that the first three characteristics enumerated above are archaisms, the remainder examples of independent phonetic change.

The phonetic structure of this language was in certain respects markedly different from that of contemporary Mongolian, and when words were borrowed from it the following changes were made, as and when required, to adapt them to Mongolian habits of pronunciation:—

(1) If a Turkish word ended in -C or -CC, a vowel, sometimes followed by an unstable *-n* or a Mongolian suffix like *-sun/-sün*, was attached to it;

(2) if a Turkish word contained a medial CC foreign to Mongolian (see Chapter X (4)), an euphonic vowel was inserted between the two Cs, and compensatory changes were sometimes made in the vocalization of the rest of the word;

(3) Turkish **ı** was changed to *i*.

(4) Turkish **p** and **v** were changed to *b* or *ğ/h* (but see (7)).

(5) Turkish **d̲-** was changed to *d* (but see (6)).

(6) Turkish **tı/ti** was changed to *çi* and **dı/di** (including **d̲ı-/d̲i-**) to *ci*.

(7) Turkish intervocalic **-ğ-/-g-, -b-, -p-** and **-v-**, whether these sounds were already intervocal in Turkish or had become intervocalic when a word was borrowed and an euphonic vowel inserted in the centre (see (2)), were sometimes changed to an intervocalic hiatus. The circumstances in which such changes were made are obscure; if the suggestion made in Chapter X (2) that intervocalic fricative *-ğ-/-g-* in pre-thirteenth century Mongolian had by that date become an intervocalic hiatus is correct then the position is that any intervocalic fricative became a hiatus.

(8) Turkish ŋ was changed to *ŋğ/ŋg*, which was followed by an euphonic vowel if it was the first component of a CC. It should however be noted that in some early texts, for example in hP'ags-pa, the original Turkish spelling was retained in words like *déŋri* (**teŋri:**).

(9) An unstable *-n* or a suffix like *-sun/-sün* was sometimes attached to a Turkish word ending with a vowel.

There are a few loan words in this lowest layer in which two or three of these changes have been combined in a very puzzling fashion; for example Turkish **idi:** has become *écen*. If **idi:** had already become ***ede:** (see (9) in the first list), it should have become **eden* (the *-n* being attached in accordance with (9) in the second list); but if the word was originally borrowed as **icin* it is difficult to see why it should later have become *écen*.

The second layer of Turkish loan words in Mongolian contains words which were borrowed, I have suggested between the eighth and twelfth centuries, from a language which had made the sound change initial **y-** > **c-**, but has not survived in any texts known to us. Kāşğarī (*I*, 31; *II*, 314) mentions the existence of such languages in the eleventh century. The languages which he actually mentions, Oğuz, Kıpçak and Türkmen, were all in the western group, and his statement was obviously too sweeping, since none of the early texts and only a minority of the modern languages in these groups show this phenomenon. But there must have been such a language in the north east at this period (perhaps Ottuz Tatar, since Kazan Tatar is one of the modern languages concerned) in order to explain not only these loan words but also the fact that this sound change is characteristic of some of the modern north-eastern languages.

While the only demonstrable difference between the language from which these words were borrowed and the contemporary " standard " languages is this sound change, there are reasons for supposing that in other respects also it differed somewhat from those languages. Some words

which for other reasons, mainly semantic, seem to belong to this layer, while clearly derived from Turkish roots, have forms which are different from those in any " standard " language.

When these words were borrowed the Mongols made certain phonetic changes in them to adapt them to their habits of pronunciation:—

(1) Turkish **ş** and **z**, which did not exist in the language from which the words in the first layer were borrowed since they had become **l** and **r**, were both changed to *s*.

(2) While a vowel, possibly followed by an unstable *-n* or a suffix, was always attached to words ending in -CC, it seems doubtful whether such a vowel was always attached to a word ending in -C. In some cases where such a -C was an unvoiced sound it seems merely to have been voiced. There are, however, several words with an attached vowel which might belong either to the first or to the second layer.

(3) The sound changes enumerated in (2) to (9) of the preceding list were made in this period also.

The third layer of Turkish loan words in Mongolian contains the words which were borrowed late in the twelfth or early in the thirteenth century after the Mongolian expansion had started. These words really fall into two different classes. The first comprises the technical Buddhist terms, mainly loan words in Turkish itself, which entered Mongolian when the Buddhist scriptures were translated into that language. Mongolian Buddhists no doubt tried to pronounce these words as they had been pronounced in Turkish, and generally speaking the Turkish spellings were retained, but it is probable that **p** and **v** were pronounced *b*, **d̲** *d*, **z**, *s* and so on. The second class comprises the everyday words, Chinese, Turkish, Persian and so on, which the Mongols picked up in the course of their conquests. Here again the natural limitations imposed by Mongolian pronunciation habits had their effect, but some

of the spellings in the *HIiy* suggest that final -ş in Turkish and other loan words was no longer changed to -*s*, although perhaps an -*i* was attached to it, and it is not at all certain how such Persian words as *bāzār* and *bāzārgān*, both of which occur in the *HIiy* (the latter in Part 2), were pronounced.

When a Turkish loan word has no phonetic characteristics which make it possible to attribute it to one of the three layers, there is no way of determining when it entered the language, unless its context or meaning provides an indication.

I shall now review the vocabulary of the *HIiy*, taking separately the words relating to particular subjects and indicating in each case which Chinese words are translated by native Mongolian words and which by loan words. The seventeen Sections of the *HIiy* provide a convenient list of subjects, but some Sections deal with more than one subject, and words relating to a subject like animal husbandry are found in two or three Sections, so a certain amount of rearrangement is necessary. In dealing with some subjects I have added a few words taken from other contemporary authorities. I have not attempted to make the lists complete, for example the glosses to the *Muqaddimatu'l-adab* contain many Turkish and other loan words which the author probably used only because he did not know the right Mongolian one, and I have drawn freely on Poppe's article *The Turkic Loan Words in Middle Mongolian*, Central Asiatic Journal I, 1. The lists of loan words differ a little from those in *The Turkish elements in 14th Century Mongolian*, as I have found one or two more since that paper was written; there are probably still more to be found.

Some of the Sections contain practically no foreign words, and it will be instructive to deal with them first, so as to show what parts of the Mongolian vocabulary were almost free from foreign elements down to the fourteenth century. The titles of the various Sections are translations of those in the *HIiy*.

SECTION 1. ASTRONOMY

The Chinese words translated by native words are:—sun, moon, star, wind, cloud, smoke, ice, snow, hail, thunder, lightning, rain, prolonged rain, fog, rainbow, dew.

There are only three words in the Section with Turkish translations:—*t'ien* (Giles 11,208), a word with a good many shades of meaning in Chinese, including " the visible sky " and " heaven " regarded as a sort of deity, is translated *teŋgiri* (**teŋri:** with the sound change **ŋ** > *ŋg* and an inserted euphonic vowel). There is a native word for " sky " *oğtarğuy*, attested in the dictionaries, but apparently not in the thirteenth and fourteenth century texts. The translators presumably preferred to use *teŋgiri* here because it had the same non-physical overtones; (2) " hoar-frost " is translated *kira'u* (**kırağu:**); (3) " the Milky Way " is translated *teŋgiri-yin oyalar* " needlework in the sky." The latter word is identified as Turkish by the plural suffix *-lar*. The word itself is derived from Turkish **oy-**, which means *inter alia* " to embroider "; the actual form ***oya** does not seem to be attested but would be a normal deverbal noun (with **-a** for the " standard " **-uğ**) in the language from which the first layer loan words were borrowed.

SECTION 2. GEOGRAPHY

This Section deals with two subjects, natural features and works of man, with a rather indefinite dividing line between them.

(1) The Chinese words for natural features translated by native words are:—two words for " earth," *ti* (Giles 10,956) translated *ğacar*, and *t'u* (Giles 12,099) translated *şira'u*,* mountain, forest, river, brook, lake, pond, dust, mud, water, waves, flood, sand-bank, bank, stone, desert, mountain pass, dry ground, *kuan* (Giles 6,368, a word with

* Both Chinese words are rather indefinite in meaning, *ti* meaning " earth (as opposed to " heaven "), ground, soil," and *t'u* " earth (as one of the five elements), the earth, ground, soil " and so on. In the Chinese translation of the *Secret History* *ti* is used to translate both *ğacar* and *etügen/ötögen*; in the *Muqaddimat* the equivalent Turkish words are **yer** for *ğacar* and **toprak** for *şira'u*.

several meanings, here perhaps " gorge "), and the mouth of a pass.

There are a good many other native words for natural features in the early authorities, for example in the *Secret History* *ğari* " a country " (translated *pang* (Giles 8,648), in Chinese usually in a political sense), *ğarki* " the course of a river " and *ğada* " cliff " (which might be a Turkish loan word if it could be assumed that *kaya:* " rock," attested as early as the eighth century, is a later form of ***kaḏa:**).

The only loan words in this sub-section are (1) " sand " *ğumaki*, not a Turkish word as such but clearly derived from **kum** " sand "; perhaps a second layer word; (2) " spring " *bulağ* (**bulak**), second or third layer; and (3) " sea " *dalay* (**taluy** in Türkü), by its vocalization a first layer word. The Turkish word is probably itself a loan word, perhaps from Chinese. Other Turkish loan words in the *Secret History* are " the junction of two rivers "* *belçir* (**beltir**) and *çiŋgis* (**teŋiz**) in Turkish " sea," in Mongolian only a proper name.

(2) As regards the works of man, words connected with agriculture and horticulture will be mentioned under Section 4. The remaining Chinese words translated by native words are:—*li* (Giles 6,908, " fence "), wall, village, path, big road (*terge'ür*, etymologically connected with *tergen* " cart "), enclosure (*küriyen*) and grave mound.

In the *Secret History* there are also *ğoto(n)* translated by the same Chinese word as *küriyen*, and *nutuğ* translated *ying pan* (Giles 13,305, 8,620, " camping ground ").

The following Chinese words are translated by Turkish loan words:—(1) " town " *balağasun* (**balık**), a typical first layer loan word; (2) " country " (*kuo* (Giles 6,609), almost invariably in a political sense), *ulus* (**uluş**), a typical second layer loan word; (3) " well " *ğuduğ* (**kuḏuğ**), second or third layer; and (4) " market " *bazar* (or *bacar?*) (Persian *bāzār*), a typical third layer loan word. Another first

* This seems to be the meaning in Mongolian. In Turkish **beltir** means " the place at which two or more roads, rivers or mountain ridges meet or cross one another."

layer loan word in the *Secret History* is *ying* " camp " (Giles 13,305), *ayil* (**ağıl**), hardly distinguishable in meaning from *nutuğ*.

SECTION 3. SEASONS

There are practically no loan words in this Section. The Chinese words translated by native words are:—spring, summer, autumn, winter, time, year, day, night, evening, the hottest month (*fu* (Giles 3,691) translated *çilger* " great heat "), hot, warm, mild, cool, cold, very cold, freezing, dry, and now. The following Chinese words and phrases are translated by native words and phrases:—mid-day, new year's day, the last night of the year. Only the following phrases contain a Turkish element, **erte:** " early ":— " bright and early " *manağar erte*, and " formerly, of old " *beler erte*.

In quoting dates of the twelve-year animal cycle in the *Secret History* and elsewhere the word used for " year " is not the native word *hon* but *cil* (**yıl**), a second layer loan word, which suggests that this method of dating was adopted during that period.

SECTION 8. CLOTHING

There are practically no loan words in this Section. The following Chinese words and phrases are translated by native words or phrases:—" garment, overlap, fur garment, boot, felt, hemp, thick clothes, sleeve, belt of clothing, hip band (the last two both translated by the same word), belt, trousers, cushion, quilt, curtain, felt stocking, embroidery, raw silk, cloth, thread, and cotton."

Other native words used in the *Secret History* are:— " trousers " *doto'açı*, " brocade " *dardas* and perhaps " ring " (if " finger ring " is intended) *dörebçi* and " shirt " *çamça* (the last much debated).

The following words are translated by loan words:— (1) " collar " *caka* (***caka:** < **yaka:**), second layer; (2) " cotton cloth " *bös* (**bö:z**), second layer, ultimately deriven from Greek *bussos* " linen "; and perhaps (3) " woven

silk" *kib*, which looks like a second layer Chinese loan word; and (4) "brocade" *cama*, which might be a first or second layer loan word with attached *-a* from Chinese *chin* "brocade" (Giles 2,068; "Ancient Chinese" *kiəm*).

There is a relevant second layer loan word in Section 7, "needle" *cö'ün* from **yigne:** which seems to have evolved as follows:—**yigne:** > **cigne:** > ***cögne:** > *cögün* > *cö'ün*.

SECTION 9. FOOD AND DRINK

The following Chinese words are translated by native words:—boiled millet, gruel, grain spirits, soup, salt, butter (translated "yellow oil"), fat, oil, meat, roast meat, dried meat, mare's milk, taste, to taste, to mix, to cook, to carve, to eat, to be hungry, and to be satiated.

The remaining Chinese words and phrases, which relate to a more sophisticated dietary, are translated by Turkish loan words. Of these four are obviously derived from Turkish roots, but do not exist in standard Turkish and may be second layer words borrowed from a Turkish dialect which, it was suggested above, contained some aberrant forms. The words translated by loan words are:—(1) "baked loaf" *ütmeg* (**ötmek**), second or third layer; (2) "cream" *tarağ*, cf. **ta:r** "clotted cream"; (3) "dried cheese" *kurud*, cf. **kuru:-** "to be dry"; (4) "scalded cream" *a'arçi*, cf. **ağar-** "to be white"; (5) "cheese curd" *bişlağ*, cf. **bış-** "to mature, ripen"; (6) "vinegar" *şirke* (**sirke:**); and (7) "vegetable drug" *em* (**em**). It seems that (8) "camel's milk" puzzled the translators and they chose *ayirağ*, a word cognate to **ayra:n** "butter-milk."

Another loan word, "(grape) wine" *bor* (**bor**) occurs in the *Secret History*.

SECTION 12. HUMAN AFFAIRS

This Section contains 130 entries, nearly all verbs with a few nouns. The great bulk of the list is basic verbs like "to hear, to see," and a good many other native basic verbs could be collected from the *Secret History* and other

authorities, but as this is a list of Chinese verbs with Mongolian translations the verbs " to be " and " to have " do not appear in it. The only Chinese words in the list which are translated by loan words seem to be the following:—(1) *jên* " to recognize " (Giles 5,609), *tani-* (**tanı:-**); (2) *shêng/hsing* (Giles 9,887), in this context " to understand," *uğa-* (**uk-**); (3) " to dance " *böci-* (**bödi:-**); (4) " to curse " *sökö* (**sög-**); (5) *shang* " to bestow " (Giles 9,735), *soyurğa-* (**soyurğa:-**); (6) " beloved " *amurağ* (**amra:ğ** with euphonic vowel); (7) " rich " *baya(n)* (**ba:y**), first layer; (8) *ch'êng* " sincere " (Giles 766) *çeŋ*; (9) " to sweep " *şi'ür-* (**süpür-** with *-p-* replaced by intervocalic hiatus); (10) " to count " *sana-* (**sana:-**). One or two other words might possibly be translated by loan words, for example " lively, alert " *bişiğun*, morphologically, but not semantically, close to **bışığ** " mature."

Some other loan words can be found in the *Secret History* and other early authorities, including such verbs as " to become " *bol-* (**bol-**) and " to shave " *dü'il-* (***düli:- > yüli:-**).

SECTION 14. NUMERALS

The digits one to nine, the tens ten to ninety, and a hundred are all translated by native words. The higher numerals " a thousand " *miŋga(n)* (**bıŋ**), first layer, and " ten thousand " *tümen* (**tümen**) are translated by Turkish loan words, the latter itself a loan word from " Tokharian B " (Kuchaean) *tumane*. " Ten thousand times ten thousand " is translated *tüg tümen*, a Turkish-looking reduplicated form of **tümen**. It is of course quite normal for primitive peoples to borrow their higher numerals from more advanced neighbours.

Of the remaining Chinese words in the Section the following are translated by native words:—" number, how many?, many, few, a kind, only or alone, single, double."

The following are translated by Turkish loan words:—(1) " a piece " *keseg* (**keseg**); (2) " half " *carim* (***carım < yarım**), second layer; (3) " half share " *carimtuğ* (***carımtuk**

< **yarımlık**); and two words which will be mentioned under " animal husbandry " in Section 5 (ii). *Carim* is not the only Mongolian word for " half "; in the *Secret History* there is also a native word, *düli*.

SECTION 15. THE BODY

This section falls into two parts.

(i) The first sixty entries relate to the body itself, its parts and its secretions. Fifty-one of the Chinese words which it contains are translated by native words. The following Chinese words are translated by Turkish loan words:—(1) " face " *ni'ur* (***ñü:z > yü:z**); (2) " fist " *nudurğa* (***ñudruk > yudruk,** with inserted euphonic vowel); (3) " kidney " *bö'ere* (**bögür**); (4) " sole of the foot " *ula* (**ul**); (5) " mole " *meŋge* (**meŋ < beŋ**); (6) " heart " *cürüke(n)* (***cürek < yürek**); (7) " beard " *sakal* (**sakal**); (8) " waist " *bel* (**bél**); and (9) " knee-cap " *toboğ* (**tobık**).

Of these the first two are certainly, and the next three probably, first layer words; the sixth is second layer.

(ii) The remaining seventeen entries relate to physical and moral qualities. The following Chinese words are translated by native words:—bald, lame, blind, deaf, dumb, fat, life, benevolence, uprightness, reliability, good order, and virtue.

Only four words are translated by Turkish loan words:—(1) " lean " *turuğa(n)* (**turuk**); (2) " knowledge " *uğa'an* (**ukuğ** " understanding "); (3) " will, resolution " *coriğ* (***coriğ < yoriğ,** which means rather " conduct, behaviour "); (4) " unwritten law " *töre* (**törö:**); the third is a second layer word, the rest first layer.

SECTION 16. CARDINAL POINTS

All the Chinese words in this Section are translated by native words:—east, south, west, north, middle, below, above, inside, outside, in front, behind, left, right, between, edge, and bottom.

There is an alternative loan word *gerü* (**geru:**) for "behind" in the *Secret History*, probably first layer.

SECTION 17. MISCELLANEOUS

This Section contains eighty-seven entries, nearly all adjectives, which are usually arranged in pairs of contrasted opposites:—"difficult, easy; existent, non-existent" and so on. Eighty-three Chinese words are translated by native words. Only the following four are certainly translated by Turkish loan words:—(1) "hard, tough" *ğata'u* (Muqaddimat, *kata'u*) (**katığ**); (2) "pure" *ari'u(n)* (**arığ**); (3) "difficult" *berke* (**berk**, properly "firm, solid"); (4) "overturned, reversed" *tetürü* (**tetrü:**). Of these the first two are probably first layer words.

Thus it will be seen that in these ten Sections, comprising 476 entries, rather over half the total, there is hardly more than a sprinkling of foreign words, and nearly all of these are demonstrably the kind of words which a primitive people might be expected to acquire when it came into contact with a more advanced one. In the remaining seven Sections there is a rather higher proportion of loan words, but these too are of the same kind as those in the Sections already analysed.

SECTION 4. FLOWERS AND TREES

This Section, with twenty-eight entries, is in a rather chaotic order and can best be divided up into four small sub-sections:—trees and other natural vegetation, parts of trees and plants, agricultural crops and horticultural crops, with a rather uncertain dividing line between the last two, which can conveniently be expanded to include agricultural and horticultural terms from other Sections.

(i) All the Chinese words for trees and other forms of natural vegetation are translated by native words:—tree, pine, juniper, cypress, elm, willow, bamboo (translated *ğulusun* which means "reed, cane" and was probably the nearest equivalent that could be found for "bamboo"), *artemisia* (*hao* (Giles 3,871) has several meanings but the

translation *şiralcin* shows that " artemisia " was meant here), weed, and grass.

The Chinese word for " bramble, thorn bush " is translated *boro keçe'ö* of which *boro* " grey " is a Turkish first layer loan word from **bo:z**.

Other relevant native words occur in the *Secret History*:—" thicket, bushes " *buta*; " feather grass " *deresün*; and *p'êng hao* (Giles 8,902, 3,871) " crysanthemum coronarium " *ğamğa'ulsun*.

(ii) The following Chinese words for parts of trees and plants are translated by native words:—twig, leaf, root, and seed. It has been suggested that *nabçin* " leaf " is connected with Turkish **yap-** " to cover," from which the Turkish word **yapurğa:k** is derived, but *dabci(n)* " cover, lid," mentioned below, is more likely to be connected with **yap-**, which in that case would go back to an earlier form ***ḏap-** and not ***ñap-**.

" Elm bark " *burğasun* is another relevant native word in the *Secret History*.

The only Chinese word in this sub-section translated by a Turkish word is " flower " *çeçeg* (**çéçeg**).

(iii) The Chinese words for agricultural crops and terminology are almost exclusively translated by loan words. Only one Chinese word for a field crop *mi* (Giles 7,802; usually " hulled rice," but also used for other grains) is translated by a native word, *amu(n)*. The other field crops all have Turkish names:—(1) " millet " *konoğ* (**konak**); (2) " barley " *arbay* (**arpa:**); (3) " wheat " *bu'uday* (**buğda:y**, with inserted euphonic vowel and **-ğ-** changed to intervocalic hiatus); (4) " dried rice " *tuturğan* (**tuturga:n**) and (5) " beans " *burçağ* (**burçak** " peas ").

In Section 2 (ii) " ditch " is translated *subağ*, which is not demonstrably a loan word, but might be connected with **su:v** " water," and " (ploughed) field " by *tariya(n)* (**tarığ**), a first layer loan word.

In Section 7 both " plough " (*ancasun*) and " sickle " (*gadu'ar*) are translated by native words.

In Section 11 (v) the word for "cultivator, farmer" *tariyaçi* (**tarığçı:**) is a first layer loan word.

(iv) The Chinese words for horticultural crops and terminology too are almost exclusively translated by loan words. Obviously in their most primitive state the Mongols must have depended for their subsistence on food-gathering as well as animal products, but there are curiously few words for wild food-stuffs in the *HIiy*, in addition to "seed" (*hüre*) and "grain" (*amu(n)*) mentioned under (i) and (iii) above. The only names of vegetables translated by native words are *chiu* "leeks" (Giles 2,279), *ğoğosun* and *hsieh* "shallots" (Giles 4,430) *maŋgir*.

The other vegetables mentioned in the *HIiy*, apart from those listed under (iii) above, are (1) "onion" *so'oŋğina* (**so:ğan/so:ğun**); (2) "garlic" *sarimsağ* (**sarumsak**); (3) "melon" *ka'un* (**ka:ğu:n**); (4) "water-melon" *arbusa* (Persian *xarbūza*); (5) "calabash" *kabağ* (**kabak**); (6) egg-plant *badiŋğa* (Persian *bādincān*); (7) "mustard" *kiçi* (**kiçi:**); and (8) "radish" *turma* (**turma:**). Of these (1) looks like a first or second layer word; none of the others are likely to belong to the first layer and (4) and (6) must belong to the third.

As regards the tree-crops, only two words are not demonstrably foreign and even these have a foreign look. "Apricot" is translated *güilesün*, a word still current; *güil* is not like any Turkish word for "apricot" but might perhaps be Iranian. "Peach-tree" in the *Secret History* is *ğatğurasun*.

The other words in the *HIiy* are all translated by Turkish loan words:—(1) "fruit" *cemişi* (**cémiş** < **yémiş**); (2) "jujube" *çibuğan* (**çıbıka:n**); (3) "pear" *alima* (**alma:** "apple" with euphonic vowel inserted); (4) "grape" *ücüm* (?) (**üzüm**); and (5) "nut" *ci'ağ* (**cağak** < **yağak**). Of these (1) and (5) must be second layer words and (4) probably third layer.

In Section 2 (ii) "garden" is translated by the Persian word *bāğ*, third layer.

SECTION 5. BIRDS AND BEASTS

This Section can conveniently be broken up into six sub-sections:—wild and mythical animals; domesticated animals (including entries in other Sections relating to animal husbandry and the use of animals for transport); fish (including entries relating to fishing); reptiles and insects; birds; and miscellaneous words including parts of animals.

(i) The Chinese words and phrases for wild animals translated by native words are:—bear, *huang yang* (Giles 5,124, 12,842, " the goitred antelope ", *Antilope gutturosa*), wild boar, wolf, fox, deer, stag, sable, otter, yak, hedgehog or porcupine, rat, and mole.

The following Chinese words and phrases are translated by loan words:—(1) " hare " *ta'ulay* (**tavışğa:n**); (2) " dragon " *lu* (Turkish **lu:**, ultimately derived from Chinese *lung*); (3) " tiger " *barsa* (perhaps to be read *bars*) (Turkish **bars**, an Iranian loan word, cf. Persian *pārs* " cheetah "); (4) " monkey " *beçin* (Turkish **biçin**, an Iranian loan word, cf. Persian *būzīna* " monkey "); (5) " lion " *arsalan* (**arsla:n**); (6) *piao* (Giles 9,112, " tiger cat ") *ğarğulağ* (**kara: kulak** " lynx "); (7) " elephant " *ca'an* (**cağan** < **yağan**); (8) " musk deer " *ciğar* (**cıpa:r** < **yıpa:r** " musk " but not specifically " musk deer "); (9) *hsi* (Giles 4,128, " rhinoceros ") *kerse* (ultimately derived from Arabic *ḫarīṣ* " rhinoceros," but the intermediary Turkish form, if there was one, has been lost); (10) " steppe fox " *kirsa* (**karsak**).

Of these (1) is a first layer loan word with the sound changes (intervocalic hiatus for **-v-**, *-l-* for **-ş-**, elision of **-ğ**) characteristic of that period, but the final *-y* for **-n** is puzzling; (2), (3) and (4) were no doubt borrowed during the second period when the twelve-year animal cycle was introduced; (7) and (8) are second layer loan words.

There are two puzzles in this sub-section:—

(a) " Jackal " is translated *çü'e böri*, a phrase which also occurs in the *Secret History*; *böri* is obviously Turkish **böri:** " wolf " but the first word is hard to explain. In

Xākānī "jackal" is **arju:**, probably a loan word since **-j-** is not a Turkish sound, but the Mongolian name survives in Tuvan **şö:börü**, Kırğız **çö:** and Uzbek **çiya bŭri**. These may be Mongolian loan words, but the balance of probability is that the phrase was originally Turkish, although not Xākānī.

(b) "Yellow rat" is translated *cumuran*; "yellow rat" sometimes in Chinese means "marmot," and both Haenisch and Levicki take that to be the meaning here; but "marmot" in Mongolian is normally *tarbağan*, a word which occurs in the *Secret History*, but is not there translated "yellow rat." According to Giles (under No. 5,124) "yellow rat" also means "a kind of weasel, *Mustela sibirica*," and that is perhaps the meaning *cumuran* is intended to translate. In any event both *tarbağan* and *cumuran* are native words.

There are other native words for wild animals in the *Secret History* including *gendü* "a male animal" and *ğandağay* "stag" or "elk" (?) and more can be found in "Classical" Mongolian like *üyeŋ* "ermine," *ergis* "male sable" and *keremü* and *burulcığana* both "squirrel." On the other hand there are also in the *Secret History* loan words like "wild game" *aba* (**a:v**), a first layer word, and "wild ass" *ğulan* (**kula:n**). In Section 11 (v) "hunter" *abaçi* (**a:vçı:**) is a first period loan word.

(ii) The Chinese words for domesticated animals which are translated by native words are:—horse, mare, foal; pig, sow; cat; dog. There are other native words connected with horses in the *Secret History*, including a generic term for "horse" *adu'un*; "a two-year-old horse" *da'ağan*; "sterile (mare)" *eremüg*; "a mare that has not foaled for some years" *esgel*; and some words for the colours of horses' coats like *ğula* "brown" and *ğali'un* "brown with a black mane and tail."

All the Chinese words in Section 7 connected with the use of horses for riding and traction are translated by

native words:—cart, wheel, shaft,* spoke, felly, hub, cart-body, bridle, reins, whip, stirrup. Other native words of the same kind occur in the *Secret History*, including "cart-body" *alan*; linch-pin *çi'un* and "saddle" *eme'el*. Another native word is "horse-herd" *adu'uçi* in Section 11 (v).

But even in this part of the vocabulary there are some loan words like "puppy" *güçüg* (**küçüg**), in Section 14 "a pair (of animals)" *ğoşi* (**koş**) and "flock, herd" *sürüg* (**sürüg**) and in the *Secret History* "crupper" *kudurğa* (**kudruğ**). In addition some of the terms connected with horse-breeding are loan words, including "stallion" *acirğa* (**adğır**), a first layer loan word, and "gelding" *axta*, a Persian word, past passive participle from *axtan* "to geld," not noted in Turkish until after the thirteenth century and so probably a third layer loan word.

In Section 11 (v) "groom" *ulaçi* (**ulağçı:**) is a loan word and so too are some of the words for the colours of horses' coats including "chestnut" *ğoŋğor* (**koŋor**) and "grey" *boro* (**bo:z**), the latter a first layer loan word.

Almost every word connected with the other branches of animal husbandry is a loan word.

(a) *Camels*. "Camel" is *temeyen* (**tevey**), a first layer loan word. Other relevant words in the *Secret History* are:—"camel stallion", *bu'ura* (**buğra:**) and "camel foal" *botoğan* (**boto:**) and in "Classical" Mongolian "gelded camel" *atan* (**atan**), "female camel" *iŋgen* (**inge:n**), and "two-year-old camel" *torom* (**torom**).

(b) *Cattle*. The only possible native word is "calf" *tuğul*, and even this may be taken from **toklı:** "six-month-old goat" with a slight change of form and meaning. The remaining words, where they can be dated, are first layer loan words; "bull" *buğa* (**buğa:**); "cow" *üneye(n)* (**ingek**); "ox" *hüger* (**öküz**, itself an Indo-European loan

* I made a stupid mistake in the postscript to *The Turkish elements in 14th Century Mongolian*. What I should have said is:—there cannot be any connection between Mongolian *aral* "shaft" and Turkish **arış** "shaft." In this meaning the Turkish word is a fifteenth century loan word from Arabic *'arış* (with an *'ayn*), same meaning. Until that date **arış** meant only "the warp on a loom"; and although morphologically *aral* could be a first layer loan word from **ariş**, at that date there was no semantic connection between the two words.

word, probably from "Tokharian B" *okso*); in Section 11 (v) "ox-herd" *hügeçi* (***öküzçi:**); and in the *Secret History* "two-year-old calf" *bura'u* (**buza:ğu:**) and "to milk" *sa'a-* (**sa:ğ-**), a word there used of other animals besides cows.

(c) *Sheep*. The relevant words, those that can be dated being first layer loan words, are:—"ram" *ğuça* (Muqaddimat *kuça*) (**koç**); "sheep" *ğoni(n)* (Muqaddimat *konin*) (**ko:ñ**); "lamb" *ğuriğan* (Muqaddimat *kurağan*) (**kuzı:**); in Section 11 (v) "shepherd" *ğoninçi* (**ko:ñçı:**); and in the *Secret History* "wool" *nuŋğasun* (***ñuŋ** > **yuŋ**) and "to shear" *kirğa-* (**kırk-**).

(d) *Goats*. The only native word seems to be "he-goat" *uğuna*. *Ku-li* (Giles 6,226, 6,930) "black goat" is translated *ima'a(n)* (**ımğa:**). In Turkish the word means "wild goat," but in the *Secret History* an *ima'an* is milked (Haenisch s.v. *sa'a-*). In the Chinese translation of the *Secret History* the same phrase is used to translate *eşige* (**eçkü:**, a generic word for "goat").

(e) *Donkeys and Mules*. "Donkey" is *elcigen* (**eşgek**). The form is at first sight difficult to explain, but in some modern north-eastern languages the form is **eştek**, which suggests the evolution **eşgek** > ***eşdek** > ***eşdik** > ***eldik** > *elcigen*. Mule is *laosa*, from Chinese *lo-tzŭ* (Giles 7,290, 12,317) but the phrase in the *Secret History ğaçidut laosasut* "mules" suggests that the Mongols also knew of the Turkish word **katır**.

(f) *Poultry*. "Domestic fowl" is *takiya* (**takığu:**), a first layer loan word.

(iii) *Fish and fishing*. Almost all the words used are native; in this Section "fish," in Section 7 two words for "fish net," in Section 11 (v) "fisherman," and in the *Secret History* "to fish" *elgü-*, "fish-hook" *elgü'ür*, and names of unidentified fish like *çurağa* and *ğadara*. The only foreign word seems to be "fish hook" *geogi*, no doubt a Chinese phrase, perhaps *kou ko* (Gilés 6,138, 6,029) "fish-hook" used in hendiadys with *elgü'ür* in the *Secret History*.

(iv) *Reptiles and insects.* The following Chinese words are translated by native words:—snake, frog, spider, butterfly, fly, gnat, worm, ant and louse; or by native phrases:—tortoise, turtle and glow-worm. The only words translated by loan words are " locust " (or " grasshopper ") *çe'örge* (**çegürge:**) and " flea " *bürge* (**bürge:**). " Gad-fly " is *köke teri'ün* " blue head," of which *köke* (**kök**) is a loan word.

(v) *Birds.* The *HIiy* contains thirty-three names of birds, obviously selected from the Chinese rather than the Mongolian repertory. It is difficult for us to identify some of them precisely, and obviously the translators themselves had difficulty in finding the exact equivalents of some of them in Mongolian.

The following Chinese words and phrases are translated by native words:—bird, black eagle, turtle-dove, *shun* (Giles 10,412, " quail "), *an shun* (Giles 52, 10,142, also " quail "), goshawk, kite, goose, mandarin duck, magpie, pheasant, duck, sparrow, *t'u hu* (Giles 12,122, 4,998, " hare-hunting falcon "), buzzard, harpy eagle, *pai hsiung* (Giles 8,560, 4,699; literally " a hundred cocks," Haenisch does not translate, Lewicki translates " vulture, hawk "; the Mongolian word *cağalmay* means " cross-wise " and in " Classical " Mongolian occurs in the phrases *cağalmay bilcuukay* " cross-bill, *Loxia* " and *cağalmay şiŋkor* " sparrow hawk, *Accipiter nisus* "), and cormorant.

Ya (Giles 12,808, " white-necked crow ") is translated *alağ ta'un* and *hu* (Giles 4,998, " a migratory bird larger than, but resembling, a crested lark; also a falcon "), *şira şiba'un*; *alağ* " dappled " is probably, and *şira* " yellow " certainly, a Turkish loan word, but the phrases were probably invented by the translators.

The " domestic fowl " *takiya* (**takığu:**) has already been mentioned under (ii).

Four names of sporting birds are translated by Turkish loan words:—(1) " sparrow hawk " *kirğuy* (**kırguy**); (2) *hai ch'ing* (Giles 3,767, 2,184, literally " sea green ") *şiŋğor*

(**soŋkur** " falcon "); (3) " peregrine falcon " *laçin* (**la:çın**); (4) *lung to êrh* (Giles 7,479, 11,325, 3,333, literally " dragon cluster "; Haenisch translates " young sparrow hawk," Lewicki " bird of prey "), *turimtay* (**turumtay** " the name of a bird of prey " (Kāşğarī)). Although this word, on which Haenisch has a long and not very convincing note, is as old as Xākānī in Turkish, it looks like a Mongolian adjective in *-tay*, and may have been borrowed by the Turks at an early date from some Mongolian tribe in the eastern steppes.

Other words are translated by loan words:—(1) " swallow " *ğariyaça* (" Classical " *kariyaça*), (**karğıla:ç/karlığa:ç**); (2) " wood pigeon " *kökörçigen* (**kökörçgü:n**); (3) " parrot " *toti* (Persian *tūtī*); (4) " peacock " *tao'us* (Arabic, and Persian *ṭā'ūs*); (5) " phoenix " *ğarudi* (Sanskrit *garuḍa*). Of these (1) looks like a first layer word; (3) and (4) were very likely acquired direct from Persian; (5) no doubt reached the Mongols through the Turks in a Buddhist context.

There are three doubtful words:—(1) " swan " *ğun*; Haenisch suggests that this is the Chinese word *hung* (Giles 5,269), but *-ŋ* > *-n* is not a probable sound change, and the word may well be a first layer crasis of Turkish **koğu:**; (2) *sung êrh* (Giles 10,449, 3,333, which Haenisch translates " young harpy eagle " and Lewicki omits) is translated *lağ*; this must be a loan word, presumably Chinese, but there is no other trace of it, *lağ* in " Classical " Mongolian means " mud "; (3) *Lao wu* which would normally mean " crow " or " rook " in Chinese is translated *tura'un*, a word used in the *Muqaddimat* and " Classical " Mongolian for " crane." It is obviously a loan word from Turkish **turña:** " crane "; the form suggests that it is a first layer loan word.

(vi) *Miscellaneous words, including parts of animals.* These are almost entirely native. The Chinese words translated by native words are:—feather, wing, claw, horn, hide, beak, tail, fish scales, navel, bone, egg, to fly, to jump, to whinny, to bark, to roar, to mew, and to fight with the horns.

21/6

The only word certainly translated by a Turkish loan word is "mane" *del* (***del** > **yel**), a first layer word; *turu'un* "hoof" has a rather Turkish look and might represent **tırnak**, but the resemblance is not close; there is in the *Secret History* another word *şiyira* for "hoof," which is certainly native.

SECTION 6. DWELLINGS

The following Chinese words are translated by native words:—house (*ger*), doorway, door, threshold, post, to open, and to shut. The Chinese phrases *wu chi* (Giles 12,737, 893) and *wu ch'uan* (Giles 12,737, 2,743) meaning "ridge pole" and "rafter" are translated literally "the backbone of a house" and "the ribs of a house."

The following words are translated by loan words:—(1) "bridge" *ke'ürge* (**köprüğ**); (2) "court-yard" *ğoriyan* (**korığ**); (3) "roof tile" *çağurasun* (**çağruk** "pounded hard by horses' hooves" in Kāşğarī); (4) "palace" *ordo ger* (the first word **ordu:**); (5) "shop" *kebid* (**kebit**); (6) "brick" *kerbişi* (**kerpiç**); (7) "pagoda" *suburğan* (**supurğan**); and (8) *ssŭ miao* "temple" *süme ger* (in which the first word seems to be a reproduction of the Chinese phrase). Of these the first three seem to be first layer loan words, and the last three probably third layer loan words.

SECTION 7. IMPLEMENTS AND UTENSILS

The words connected with agriculture have already been mentioned under Section 4 (iii), and those connected with vehicles, harness and fishing under Section 5 (ii) and (iii). The remainder can be divided up into three sub-sections:—weapons, household goods and furniture, and miscellaneous.

(i) *Weapons.* The following Chinese words are translated by native words:—lance, sword, axe, bow, bowstring, arrow, armour, helmet, and small drum. *P'ao* (Giles 8,742, "balista" and from the thirteenth century onwards "cannon") is translated by *orbu'ur*, no doubt a native word; Lewicki suggests, probably correctly, that it is connected

with *ğarbu-* " to shoot (an arrow)," which occurs in the *Secret History*. Another native word used in the *Secret History* is *ğor* (*kor* in *ibn Muhanna*) " a quiver."

The following Chinese words are translated by loan words:—(1) " big drum " *körge* (**kövrüg**), no doubt a first layer word; (2) " shield " *ğalğa* (**kalkan**) and " banner " *oraŋğa* (**oruŋu:**). In Section 11 (v) " soldier, or troops " is translated *çerig* (**çérig**) and " warrior " by *ba'atur* (**bağa:tu:r**).

There are several other Turkish military terms in the *Secret History* including " outstanding warrior " *bökö* (**böke:**), " detachment of troops " *böleg* (**bölük**) and " reserve troops " *geçige* (from **kéç-** (**géç-**) " to hang back "). They have a rather second layer aspect.

(ii) *Household goods and furniture*. I include in this subsection articles which may have been used in stables, etc., since there is no means of distinguishing them.

The following Chinese words are translated by native words:—trough, cord, plate, spoon, wooden dish, sieve, cupboard, sack, comb, walking stick, vessel or container, kettle, pestle, mortar, table, mirror, mat, bed, head-rest, lock, key. There are other native words of the same class in the *Secret History*:—" spade " *çalir*; " rope " *arğamçi*; " birch-bark bucket " *dağtay*; and " leather bucket " *gö'ür*.

The remaining words are translated by loan words:—(1) " bottle " *luŋğa*, its initial *l-* excludes the possibility that it is a native word, probably a Chinese loan word; (2) " bowl " *ayağa* (**ayak**); (3) " cup " *çağun ayağa*, probably " a bowl on a foot-stand " (" Classical " Mongolian *çaku(n)*); (4) " long-handled pan " *irağa*; Lewicki suggests that this is Turkish **ırğağ** " hook "; this is phonetically possible, especially if this is a first layer word; (5) " bamboo curtain or screen " *çiğ* (**çı:ğ**); (6) " lamp " *cula* (**cula:** < **yula:**); (7) " lamp bowl " *culabçi*, the same with a Mongolian suffix; (8) " ladder " *geçgi'ür*, not as such a Turkish word, but probably derived from **keç-** (**geç-**) " to cross, pass

over"; (9) "broom" *şi'ürge* (*süpürgü:*), the verb *şi'ür-* (**süpür-**) occurs in Section 12; (10) "weighing machine, scales" *batman* (**batma:n** which means "a specific weight," originally about a kilogramme); (11) "scissors" *ğayiçi*, not as such a Turkish word, but derived from **kıy-** "to cut up"; (12) "chair, couch" *sancali* (Persian *sandalī*). Of these (4) and (9) look like first layer loan words; (6), and consequently (7), are certainly, and (8) and (11) probably, second layer loan words, and (12) is certainly third layer. There are one or two other loan words of this class in the *Secret History*; "beaker" *çuŋ* is no doubt Chinese *ch'ung* (Giles, 2,891), same meaning; and "the cover (of a pot, quiver, *etc.*)" *dabçi(n)* seems to be connected with Turkish **yap-** (***d̲ap-**?) "to cover," if so a first layer word.

(iii) *Miscellaneous*. Several words in this sub-section are connected with writing. The only word which is not demonstrably a loan word is "paper" *ça'alsun*. All the rest are translated by loan words:—(1) "writing brush" *ücüg*, Turkish **üjek** "a written character, an early loan word from Chinese *tzŭ* (Giles 12,324, "Ancient Chinese" *dz'ieg*), same meaning; (2) "ink" *beke* from Chinese *mo* (Giles 8,022, "Ancient Chinese" *mek* (probably in fact *mbek*); (3) "seal" *tamğa* (**tamğa:**). The *Secret History* contains several other relevant words:—"to write" *biçi-* (**biti:-** ultimately derived from Chinese *pi* (Giles 8,979, "Ancient Chinese" *piĕt*) "a writing brush"); "letter" *biçig* (**bitig**); and "book" *debter* (Persian *daftar*). The last must be a third layer loan word, the others cannot be firmly dated.

The remaining miscellaneous words translated by native words are:—stringed instrument, fire, and ash. "Gong" *çaŋ* might be a mere onomatopoeic, but the same word with similar meanings occurs in Turkish and Persian. "Castanets" *çargi* is Turkish **çalğı:**.

SECTION 10. PRECIOUS THINGS

This is a short but very significant Section. The only Chinese words translated by native words are:—(1)

" silver " *müŋgü(n)*; (2) " copper " *ces*; (3) " copper ore " *şiremün*; (4) " tin " *tu'ulğan*; and (5) " pearl " *subud*, which is not demonstrably a loan word. *Shui yin* (Giles 10,128, 13,253) " quicksilver " (literally " liquid silver ") is translated *göleyen usun* " shining water," which looks more like a phrase invented by the translators than a native phrase. " Copper cash " is translated *co'os*, which looks like a Mongolian rendering of a second period loan word like ***coğoş** < ***yoğoş**. No such word exists in Turkish, but it might be cognate to **yoğun** " thick."

The following words are translated by loan words:—(1) " gold " *altan* (**altu:n**); (2) " iron " *temür* (**temür**); (3) " jade " *ğaşi* (**kaş**); (4) " jewel " *erdini* (**erdini:**, ultimately derived from Sanskrit *ratna*); (5) " large pearl " *tana* (Persian *tāna*); (6) " crystal " *bolor* (Persian *bulūr*). Of these (1) must be a first layer loan word; (3) with its *-şi* cannot be earlier than the second layer and is probably third; (4), which belongs to Buddhist terminology, (5) and (6) must be third layer loan words.

In the *Secret History* we have " steel " *ğata(n)*, which seems to be an alternative form of *ğata'u* (Muqaddimat *kata'u*), " hard, tough," a first layer loan word from **katığ**, and " blacksmith's bellows " *gürege/kü'ürge* (**körük**).

SECTION 11. MEN AND THINGS

This is a long and miscellaneous Section containing eighty-seven entries. Some of them have already been mentioned in Sections 4 (iii), 5 (ii) and 7 (i). The remainder can conveniently be broken up into five sub-sections:—political; religious; terms of relationship; personal pronouns; and professional and miscellaneous.

(i) *Political.* There are only one or two political terms in this Section, but they provide a convenient opportunity for assembling the words relating to political organization which occur in the *HIiy* and some other early authorities. By the fourteenth century when the *HIiy* was compiled, the political structure of the Mongolian Empire was fully evolved and it can be taken as certain that most of the

words concerned had acquired new and extended meanings. Moreover, since the *HIiy* is a list of Chinese words with Mongolian equivalents, it cannot be expected to throw much light on the true nature of Mongolian institutions, particularly in early periods, but it is clear that the Mongols got most of their political terminology from the Turks. The only words in this Section translated by native words are "official" *noyan* and "minister" *tuşimel* (a deverbal noun from *tüşi-* "to rely on"). Other native words found in the *Secret History* are "governor (of a town or province)" *daruğa*, and "tax" *ğubçir*.

The following words are translated by Turkish loan words:—in this Section "Emperor" *ğağan* (*Secret History* *ğağan/ğa'an/ğan*; hP'ags-pa *ğa:n*) (**kağan**); "ambassador" *elçi(n)* (**élçi:**); "archivist" *biçeçi* (**bitigçi:**), in Section 6 "palace" *ordo ger* (**ordu: ger**), in Section 15 *li* (Giles 6,949) "ceremony, ritual" *töre* (**törö:** "unwritten law," which is no doubt the meaning of the Mongolian word), and in the *Secret History* "palace" *ğarşi* ("Classical" *karşi*) (**karşı:**).

There is in the *Secret History* a block of second layer loan words relating to administration, some of them in forms which do not actually occur in Türkü or Xākānī but are undoubtedly Turkish by origin. These include "posting station" *cam* (**cam** <, **yam**); "post rider" *camçi* (**camçı:** < **yamçı:**); "legal proceedings" *carğu* (**carğu:** < **yarğu:**); "order, ordinance" *carliğ* (**carlığ** < **yarlığ**); "to organize, put in order" *casa-* (**casa:-** < **yasa:-**), and "law, order" *casağ* (**casa:ğ** < **yasa:ğ**).

There are also in this Section three names of nationalities, all reflecting the terminology of the fourteenth century:—*han jên* "Chinese" translated *Kita*; *ta ta* "Tatar" translated *Maŋğol*; and *hui hui* (originally "Uyğur," at this period "Moslem") translated *Sarta'ul* (**sart**, the Turkish form of Sanskrit *sarṭha* "merchant," with a Mongolian suffix).

(ii) *Religion.* There are only four words relating to religion translated by native words:—(1) "devil" *çitkör*;

(2) " ghost " *oŋğon*; (3) " female shaman " *iduğan*; and (4) " fortune teller " *tölegeçi*. *İduğan* seems to have been a Kitan word and occurs in an account of hostilities between the Türkü and the Kitan in the middle of the eighth century, see A. E. Dien *A possible early occurrence of Altaic* iduğan, Central Asiatic Journal II, 1. *Tölegeçi* is derived from *tölege*, which is translated in the *Secret History* by *kua* (Giles 6,311), the word used, *inter alia*, for the " diagrams " in the *I Ching*.

There are in this Section one first layer loan word " male shaman " *bö'e* (**bögü:**) and three words taken from the extensive Turkish terminology of Buddhism, mainly Chinese by origin:—" teacher " *bağşi* (**baxşı:**); " Buddha " *burxan* (**burxan**); and " monk " *toyin* (**toyın**).

(iii) *Terms of relationship*. All the peoples of eastern Asia had an elaborate system of terms of relationship, which reflected their family and clan organization. As K. Grønbech pointed out in *The Turkish system of kinship*, Studia Orientalia Joanni Pedersen . . . Dicata, Copenhagen, 1953, while most peoples think in terms of generations, one's own, one's father's, one's children's and so on, the early Turks thought more in terms of age groups, and so, for example, used the same word for " one's elder brother " and " one's father's younger brother." The Chinese thought in terms of generations, and the Chinese words in this Section reflect that idea; it is possible that the Mongols too thought in terms of generations, but they may have had the same ideas as the Turks since the same word is used for " younger brother's wife " and " son's wife." Nevertheless the actual Mongolian words are completely different from the Turkish.

The following Chinese words and phrases are translated by native words and phrases:—ancestors (*üridüs*, literally " men of old "), great-great-grandfather, great-grandfather, *kung kung* (Giles 6,568, 6,568; normally " husband's father," but the translation *ebüge* and the position in the list suggests that it here means " father's father "), *niang niang*

(Giles 8,242, 8,242, "father's mother"), father, father's elder brother, father's younger brother, father's sister (*ağay egeçi* "elder brother's elder sister"), mother, mother's brother, mother's sister (*nağaçu egeçi* "mother's brother's elder sister), elder brother, younger brother, elder sister, younger sister, elder brother's wife, younger brother's wife (see above), wife, (male) child, female child, *tzŭ* (Giles 12,317, "offspring, either male or female"), son's wife (see above), daughter's husband, *chih* (Giles 1,819, "brother's son or daughter"), wife's sister's husband, *ch'in chia* (Giles 2,081, 1,139, "relations by marriage"), *ch'in chüan* (Giles 2,081, 3,145, "relations with different surnames"), adopted son (*tece'emel kö'ün* "foster child," probably a phrase invented by the translators).

The Turkish (originally Sogdian?) word **xatun** is used to translate *niang tzŭ* (Giles 8,241, 12,317), which seems to have been taken as an honorific term for "wife," since this word, and not *gergey*, the usual word, is used in the phrases translating "wife's father" and "wife's mother." Only two other Turkish loan words occur in the list; "sister's child" *ceye* (***cegen** < **yegen**), a second layer word, and "grand-child" *açi* (**atı:**). Grønbech, *op. cit.*, pointed out that the first originally meant "son of one's younger sister or daughter" and the second "son of one's younger brother or son."

There are also in this Section some more general terms:—old people, old woman (translated by the same word as "father's mother"), widow, small child, male and female. The only foreign word is in "male" *ere gü'ün* (**er**), no doubt a first layer word. *Ere* was also used for "husband," since there is no native word for this, or at any rate none has survived.

(iv) *Personal pronouns.* Six are listed:—"I, thou, we, my, thy, his," all native words. There is also "self" *ö'erü(n)*; it has been plausibly suggested that this is a first layer loan word from **ö:z**.

(v) *Professions and miscellaneous.* The following Chinese

words describing professions are translated by native words:—singer, leather worker, and cook; but "craftsman" *ura(n)* (**u:z**) and "physician" *otoçi* (**ota:çı:**) are loan words, the first from the first layer.

Of the remaining miscellaneous words the following are translated by native words:—"man, people, friend, guest, slave, thief."

Nökör is given as the equivalent of *pan tang* (Giles 8,603, 10,721). Haenisch and Lewicki both translate this phrase as "companion," but Giles' translation is "the servant of a petty military officer," and this is likely to be nearer to the true meaning since as a loan word in Persian *nawkar* means specifically "servant, employee."

There is also one first layer loan word "master" *écen* (**idi:**), the phonetics of which were discussed above.

SECTION 13. SOUND AND COLOUR

There are eleven words for various colours in this Section, and other words, native and foreign, for the colours of horses' coats were mentioned in Section 5 (ii). The Chinese words translated by native words are curiously few:—red, white, green, and plain coloured. The words translated by loan words are:—(1) "colour" *öŋgö* (**öŋ**); (2) "yellow" *şira* (**sarığ**); (3) "grey" *boro* (**bo:z**); (4) "camel coloured" *temeyen öŋgö* (**tevey öŋ**); (5) "purple" *cihi'in* (**yipkin**); (6) "blue" *kökö* (**kök**); (7) "black" *ğara* (*Muqaddimat kara*) (**kara:**); and (8) "crimson" *al* (**a:l**).

Of the remaining words included rather arbitrarily in the Section, native words are used to translate:—sound, shadow, brightness, and air. "Incense" is translated by *güci* (Turkish **küji:**, which is probably a loan word, perhaps Sogdian).

This analysis of the vocabulary of thirteenth and fourteenth century Mongolian, and the list which it discloses of the subjects about which the primitive Mongols could talk almost without being compelled to use a single foreign word:—astronomy, geography, the seasons, clothing, food and drink, human affairs (i.e. verbs), numerals, the body

and its parts, the cardinal points, miscellaneous matters (i.e. adjectives); within the remaining sections:—trees and natural vegetation, parts of trees and plants, horses and horse transport, pigs, dogs and cats, fish and fishing, reptiles and insects, parts of animals, weapons, terms of relationship and pronouns, and, at a primitive level, dwellings, household goods and religious ideas, make it clear that the native vocabulary by itself was perfectly adequate for the needs of a Bronze Age people living in small groups in the Siberian forests. Indeed an anthropologist by a careful study of these native elements could reconstruct in broad outline the way in which the primitive Mongols lived. As I pointed out in *The Turkish elements in 14th Century Mongolian*, the early accounts of the Kitan in the Chinese histories describe a people on about this level of development. Conversely the list of subjects regarding which there is a high concentration of loan words:—agricultural and horticultural crops, most domestic animals, buildings, writing and politics, indicates the fields in which these primitive Mongols looked to their more advanced neighbours the Turks for inspiration and guidance. Indeed the tentative classification of these loan words in three layers makes it possible, obviously in no more than a broad outline, to estimate in what directions advances were made in the three periods corresponding to these three layers. Thus, for example, the appearance in the first period of loan words for " town," " bridge," " craftsman," " locust " and steppe animals like " the hare " points to a transition from primitive life in the forests to more organized life in the open country. The main advances in animal husbandry and agriculture seem to have occurred during this period. The second period seems to have been marked by an advance in horticulture, improvements of clothing (" collar ") and household equipment (" lamp "), and a great development in political institutions, including the adoption of the twelve-year animal cycle. The chief features of the third period were perhaps the introduction of Buddhism and a great broadening of contacts with the

outer world typified by the adoption of Persian words for " market," " merchant," exotic animals like " parrot " and " peacock " and exotic fruits like " egg-plant " and " water melon."

EPILOGUE

AN OLD-FASHIONED LOOK AT THE LINGUISTS

Dear Sir Gerard,

I write on behalf of the Editorial Committee of the Philological Society to thank you for your article. They found themselves in broad agreement with your observations, but regret that in view of other commitments it will not be possible to include it in the Society's Transactions.

(*Signed*)
Honorary Secretary for Publications.

When I was an undergraduate at Oxford before the First War, the science concerned with language was called "philology" and its practitioners "philologists." "Linguists" were chaps who were rather good at talking two or three foreign languages, often because they were of mixed racial origin, useful to have about the place on a continental tour, but somehow faintly non-U. I think that this mild disdain for the linguists was largely due to the fact that they were so insensitive to the finer points of etymology as to describe themselves by a Latin word with a Greek suffix. We used to be rather particular about such things.

We were of course intolerant and a little unfair; however dubious its etymological ancestry, *la linguistique* was a recognized science on the Continent, even if the word "linguistics" had not established itself in this country, and some of the finest work ever done in that field was being done here, though under another name. Anyhow time has brought its revenges. Those halcyon days came to an abrupt stop; the First War and the need to earn a living diverted my interest into other channels, and it was only recently that I was able to return to the old love, only to find during my prolonged bathe in worldly affairs the linguists had succeeded in stealing nearly all the philologists' clothes, and were busily occupied in getting the rest. I need only quote one of the most respected high priests of that mystery, the late Professor J. B. Firth, who in his

recent book *Papers in Linguistics* 1934-1951, page 217, footnote 1, wrote: "I think contemporary English usage would substitute *linguistics* and *linguistic* for Professor Hjelmslev's *philology and philological*."

My own, perhaps unduly old-fashioned, view is that in recent years the linguists have been getting altogether too uppity. It is bad enough in this country, but in the United States, that home of brinkmanship, they have recently been on, or even over, the brink of declaring that the philologists are a bunch of out-of-date, no-good old fogies struggling in vain against the healthy gusts of fresh air that are blowing the dust and cobwebs out of the halls of learning. One young gentleman, for example, recently took it upon himself to write what he called a "descriptive grammar" of pre-Classical Mongolian, apparently unprejudiced by any previous knowledge of the subject, and without reference to the admirable work of previous scholars, living and dead, in that field.

The purpose of this little essay is to suggest that philology and linguistics really are different, though closely related, sciences, and that there is everything to be said for defining quite clearly the sphere of each of them and the extent of common ground that lies between them. The scope of language study is a vast one; anyone who has any doubts on that subject should read Professor Firth's book just quoted; and nothing will be lost by trying to introduce some order into it. Otherwise every young linguist will feel that he has got to cover the whole ground, and end up, like the stockbroker in the story, by knowing nothing about everything.

My point, put quite shortly and based on pure etymology (*logos* "word": *lingua* "tongue"), is that the proper study of the philologist is the written word, and his particular concern the structure and history of languages, while the proper study of the linguist is speech, and his particular concern the use of sound to convey meaning. Admittedly languages were spoken for thousands of years before anyone thought of writing them down, so to that extent the raw

material of the linguist came into existence first, but language study did not begin until languages had been written down, and the raw material which was used when it did begin, perhaps not far short of four thousand years ago in Mesopotamia, was the written word. In other words philology started many centuries ago; linguistics, in my sense of the term, did not really start much before the nineteenth century.

The techniques which were invented by the earliest philologists, and they were clearly invented again and again quite independently in different places and at different times, were, to use the terms we now employ, (1) grammar with its two branches, (a) morphology, that is the study of the variations in the forms of words in different contexts, and (b) syntax, that is the study of the mutual relationships of words in the sentence; (2) lexicography, that is the compilation of lists of words and their meanings; and (3) etymology, that is the study of the reasons why certain words have the meanings that they have. In its earliest stages philology was what the linguists call "synchronic," that is it dealt with the language concerned as it then was, and its function was mainly normative, that is to lay down, or try to lay down, correct forms, usages and meanings. But language never stands still, it is changing the whole time, and before very long the philologists found that their work was getting "diachronic," that is that the language which they spoke and wrote was slightly, or not so slightly, different from that which their predecessors had studied and with which their grammars and dictionaries dealt. From that time onwards philology, without ceasing to be normative, also became historical, that is to say it began to trace the history of the language concerned through its various changes of form, usage and meaning. So far as its normative functions were concerned it was generally fighting a losing battle against natural change, but it did at any rate try to ensure that in spite of these changes the language was still used in such a way as to convey meaning precisely and not to blur or conceal it.

It is not wholly untrue to say that philology, since its raw material is the written word neatly laid out before it for dissection and analysis, is mainly concerned with dead matter. It is not primarily concerned with sounds, or the correct pronunciation of words or passages of speech, but it has never been wholly insensitive to these things. After all the meaning of some words, for example "bow" in English, is determined *inter alia* by their pronunciation. Indeed it was in the phonetic field that some of the finest work of the nineteenth century philologists was done. It was by acute phonetic analysis that the mutual relationships of the Indo-European languages were worked out and their putative common ancestor tentatively reconstructed. But that was in a sense incidental. It made no difference to the philologists whether the texts which were their raw material were direct transcripts of the spoken word or original written compositions. In fact in some texts, like certain of the Chinese Classics, the style was so compressed, and the language had reached such an advanced stage of phonetic decay, with an enormous proliferation of homophones, that the texts, if read aloud, would not even have been intelligible. Similarly there are compositions in, for example, Persian (the script of which is often highly ambiguous unless very fully provided with the vowel points customarily omitted) which were quite deliberately written in such a manner that by reading the same written words in different ways two quite different meanings could be conveyed. These are of course extreme cases, but they are sufficient to prove the fact that there are some kinds of language study which are far removed from the simple study of the use of sound to convey meaning.

Conversely, if the raw material of the philologist is in a sense dead matter, the raw material of the linguist, as I understand that term, is essentially living matter, the sounds with which men express their thoughts, wishes and the like. The most useful and productive part of his field nowadays is the study and recording of the numerous languages of Africa, America, Australasia, and even to a

lesser extent Asia which are still purely spoken languages and have never been reduced to writing. Here he is a real pioneer, and it is interesting to note that the old tradition of independent invention of techniques which has always been so characteristic of language study is still very much alive. Professor Firth, in his book already referred to, pp. 221 ff., quotes the Report of the Commission set up by the International Council for Philosophy and Humanistic Studies, Unesco, Paris, 31 May-7 June 1951, in which suggestions are made, *inter alia*, for collecting the vocabulary material of unwritten languages. It is interesting to note that the categories of words to be collected—numerals, kinship terms, parts of the body, *etc.*—read very like the chapter headings of the Chinese vocabularies of "barbarian languages," Mongolian, Turkish, Hsi hsia, *etc.*, compiled in and even before the early days of the Ming Dynasty (late fourteenth century), and those used in the Arabic vocabularies of Turkish and Mongolian compiled in Western Asia and Egypt in the thirteenth to fifteenth centuries. There is no reason to suppose that any of these groups of scholars, except perhaps the Unesco group, knew anything of the work of the others.

It would of course be quite wrong to suggest that the linguist should confine himself to the study of unwritten languages. If that were so, he would have perhaps not more than another century of useful work to look forward to. There is much work for him to do even on well-known languages, on which the philologists have, by and large, done as much as they can be expected to do, until natural change creates conditions in which they can carry the history on a little further. It is for the linguist not the philologist to frame the rules for pronunciation, accentuation, intonation and the like in the handbooks of living languages.

Nevertheless the linguist is in much the same kind of quandary as that which confronted the Royal Air Force in the early days when it was fighting for the right to an existence independent of the Army and the Navy. This

was, it will be remembered, vividly (but in actual fact more succinctly and alliteratively) described in an allegory by the distinguished Admiral who remarked that even the rosy-plumaged birds could not reproduce their species on the wing. The hard fact is that languages cannot satisfactorily be studied when they are merely spoken orally (and heard aurally), or even when they are recorded on gramophone record or magnetic tape. For scientific examination and analysis they must be written down. And a language that is written down, however elaborate the system of phonetic transcription employed and however trivial the utterances recorded, thereby immediately becomes, *ipso facto*, potential meat for the philologist. The linguist cannot of course be excluded simply by reason of the reduction of the utterance to written form; it is he, not the philologist, who must explain the different shades of meaning conveyed by stressing the second, fourth or fifth words in such a statement as "It is a nice day." Indeed it is one of the most important functions of the linguist to explore the influences of stress, accent and tone on meaning, and to discover by the use of modern techniques like palatography and kymography exactly how particular languages are pronounced. But, when all is said and done, these phonetic niceties, though important, are not of the first importance. Without grammar and lexicography to back him up the linguist cannot really get anywhere, and it would be foolish of him to push the tradition of independent invention of techniques to the point of putting everything that the philologists have done over the past three or four millennia into the waste-paper basket and start again from scratch to write "descriptive" grammars of languages for which perfectly satisfactory grammars already exist. Nor do the saner linguists have any such intention. Professor Firth (*op. cit.* p. 216) has some wise words to say about the merits of the traditional grammars of Sanskrit, Latin, Greek and Arabic. But there are angry young men about to-day in every walk of life, and linguistics is (or are?) not free from them.

My plea is for greater mutual understanding, tolerance and cooperation. Let the philologists and linguists concerned with each language get together and divide the work between them in such proportions as seem suitable in the particular circumstances of the case. Ideally for the living written languages which in this country are the lesser known ones, languages like Turkish, Mongolian, Tibetan and those of the smaller European countries (Finland, Portugal, *etc.*), where only limited man-power is available, the same scholar should equip himself both as a philologist and a linguist. Where the volume of material is so great, as it is in the case of the major European languages, Chinese, Arabic and the like, that it is beyond the capacity of one scholar to master the whole subject, then let us have an agreement that the history of the language and its grammar, lexicography and etymology, at any rate so far as the written forms of the language are concerned, are the field of the philologist, and that it is for the linguist, on the sure foundation laid by the philologist, to explore its phonetic structure and the intricacies of the spoken language, both in its oral form and when it is embodied in writing in what is called the colloquial or informal style. Heaven knows that this branch of the subject is sufficiently intricate to give the linguist enough to do, without his trying as well to do for himself what has already been done for him by the philologist better than he could do it without another lifetime of study.

One final problem remains, that of the study of "language" as opposed to languages. In this case dichotomy is less easy; nevertheless surely the same principles hold good. The philologist who is not content to confine himself to the study of a single language or group of languages is surely the right man to generalize his experience and try to discover whether there are general principles which lead men all over the world to formulate their thoughts in words in much the same kind of way by the use of a combination of nouns, verbs and supplementary words (adjectives, adverbs, etc.). By parity of

reasoning the linguist who has had great experience in recording and analysing the use of sound to convey meaning is surely the right man to compile a corpus of the sounds so used and see whether he can establish any general correlation between the two, as for example Professor Firth (*op. cit.* p. 44) thinks he has established a correlation in English between the initial sounds *sl-* and *sn-* and pejoration. Even on this rather ethereal level the subject is so vast that division of labour seems imperative if real progress is to be made.

Page 72. Since this volume went to press the Kırğız Academy of Sciences has published *Novye Epigraficheskie Nakhodki v Kirgizii*, Frunze, 1962, announcing the discovery of four new Talas inscriptions and the rediscovery of the second old one which was believed lost. It is clear that all nine inscriptions come from a single cemetery at an ancient site called Ayrtam Oy about seven kilometres east-south-east of the town of Talas on the river of that name. The cemetery is adjacent to a walled town which is datable by pottery and other small finds to the early Karakhanid period, tenth-eleventh century, but it is said to be older than that town, though there does not seem to be any good reason why it should be much older. The whole area has been cultivated for many years, the stones have been moved, the grave mounds and earth walls on the site have been largely levelled, and it seems impossible to associate any of the stones with individual graves. The book contains a hand copy of the second old inscription better than that in *Malov*, 1959, page 60, and hand copies of three of the new inscriptions; one is only provisional since this and the fourth stone require extensive cleaning. No continuous sense can be made out of any of the inscriptions; indeed there are almost no finite verbs and no clear links between the individual lines in the inscriptions. There is one recurring phrase; **otu:z oğlan sağdı:çları:**, " thirty youths, his close friends," in line 1 of the second old inscription (which seems to begin with **er atım** in line 2) reappears (but with **sagdı:çı:**) in two of the new inscriptions and, somewhat distorted, in the fourth old one. This suggests that these stones are roughly contemporary. In the eighth century this area was probably within the dominions of the Western Türkü (see *Ak Beshim-Suyab*), and these stones were perhaps originally placed on the graves of petty chieftains in that confederation.

INDEX

An asterisk against a page number indicates that the full title of a book or article by the author named is mentioned on that page. To that extent the index serves as a rudimentary bibliography.

Aalto, P. 68*.
'Abdullah, ruler of Golden Horde 183.
ain 76.
'alāma 176.
aleph 75.
alphabets See Arabic *etc.*
Altaic theory II, V, VII, 36, 37, 215 ff.
Arabic alphabet 119 ff.
script, Mongolian texts in 192 ff.
Turkish texts in 46 ff., 62, 67, 119 ff.
Aramaic alphabet 74 ff., 101, 105, 114.
Arat, R. R. 48*, 50*, 180***, 184*, 185*.
Argun, Ilkhan 182.
Armenian authorities on Mongolian 195.
script, Kıpçak texts in 48.
A-shih-na 8, 9, 27.
Atalay, B. 50*.
Atebetü'l-Hakayik (*Atabatu'l-Ḥaqā'iq*) 48, 180, 183, 185.
ATG 68.
atım 71.
Avar people 26, 77, 167.
Azeri language 49, 161.
B- initial, in Turkish 69, 91, 130, 161, 218.
Bacot, J. 122*.
bağa:tu:r 19, 72, 88, 197, 239.
Bailey, Sir Harold 29*, 86*, 93*, 123*.
Bang, W. 184*.
Bárczi, G. 127.
Baskakov, N. A. VI*, 129*.
Bawden, C. R. IX.
baxşi: 185, 243.
Bazin, L. VI.
beg 15.
Benzing, J. 9*, 38*, 60*, 125*, 216*.
beth 75.
in M.O.A. 180, 183, 184, 187.
Bichurin, N. Ya. 7*, 28, 30.
Bilge: Kağan 68, 89.
BK 68.
Blake, R. P. 195*.
bo:d bodun 11, 148.
Bonaparte, Prince R. 182*.
Boodberg, P. 28.
Boyle, J. A. 47.
Brāhmī alphabet 67, 86, 91 ff., 106 110, 111, 162.
script, Turkish texts in 91 ff.
Brockelmann, C. 144*.
Buddhist texts in Mongolian 182, 186, 221.
Turkish 50, 51, 92, 96, 101, 108 ff.
Bulgar language 38 ff., 41, 46, 125.
people 20, 41, 127.
Burhānu'l-dīn, Qāḍī 49
Byzantine authorities, Turkish names *etc.* in 2, 6, 38, 41, 123 ff.
Greek alphabet 73, 79.
C 136.
c- initial, in Turkish 170.
-c- intervocalic, in Turkish 167.
Caferoğlu, A. 101*.
Çağatay language IV, V, 47, 49, 52 ff., 66, 184, 194.
Calālu'l-dīn al-Rūmī, Mawlānā 49.
caph 76.
Cc 136.
CC 136.
in Mongolian 196, 206 ff.
in Turkish 166 ff., 196.
CCC 136.
in Turkish 166 ff., 196.
lacking in Mongolian 207.
Chavannes, E. 7*, 8, 13, 26, 29, 30, 88.
Chêng, Tê-k'un 17*.
cheth 75.
Ch'ien Han Shu 9.
Chih-chih *shan-yü* 22, 24, 28, 39.
Chin Shu 9.
Chinese :—
Ancient language 4.
Archaic language 4.
authorities on Turkish history 2 ff., 122.
early contacts with west 17.
loan words in Mongolian 235 ff.
Turkish 15, 232 ff.
in Tibetan script VIII, X, 5, 123.
transcriptions of Mongolian 126, 191 ff.
Turkish IX, 4, 6, 13, 87 ff., 122.
Chinggis Khan 31, 179.
Chiu T'ang Shu 8.

Chou Shu 8.
Christensen, A. 113*.
Christian texts in Turkish 51, 101, 106, 175.
Chuastuanift 116, 118.
Chuvash language 38, 41, 60, 128, 130, 161, 215.
çigşi: 15, 19.
Cleaves, F. W. 181*, 195*.
Codex Cumanicus 20, 48, 52, 57, 165.
commercial texts in Uyğur script 51, 101, 106, 113.
Constantine Porphyrogennetos 124.
Cordier, H. 183*.
Czeglédi, K. 19*.

D- initial, in Turkish languages 131 160, 170.
-d- intervocalic, in early Turkish 166.
-d final, in early Turkish 166.
d- initial, in early Turkish 122, 124, 127, 160, 170.
daleth 75.
dayık 124–5.
Dien, A. E. 243*.
differentiation of letters 102.
***dıla:n** 125.
Dīwān lugāti'l-Turk. See Kāşğarl.
dogia 124.
Donner, O. 73*, 78.
Dulaurier, E. 195*.
Dzhavakov, Prince I. A. 195*.

É closed, in Mongolian 190, 204.
Turkish 83, 89, 94, 163, 205.
Eberhard, W. 4*, 7, 9, 12, 13.
Eckmann, J. 48*.
er atım 71, 72, 256.
Ergin, O. 180*.
Éştemi *kağan* 73.

F in Turkish languages 167–9.
Firth, J. B. 248* ff.
Fr. 97.
Frye, R. N. 195*.

G- initial, in Turkish languages 100, 131, 160, 171.
g and **ğ** phonetic values in early Turkish 90, 95, 167.
Gabain, A. von 68*, 78, 91* ff., 108*, 115, 117, 142, 144, 175*.
Gancak (*Kençek*) language 132.
Gandjei, T. 48*.
Gauthiot, R. 102*, 103*.
Georgian authorities on Mongolian 195.
Ghirshman, R. 80*.
Giles, H. A. 15*, 19, 223 ff.
gimel 75.
Gombocz, Z. 126*.
Greek. See Byzantine.
Groot, J. J. M. de 7*.
Grønbech, K. VII*, 68*, 81, 243*.
Grousset, R. 179*.
Grumm-Grzhimaylo, G. E. 8*.
Gumilev, A. N. 7*, 22*.
al-Ğuzz 46; see Oğuz languages.

H in Mongolian 179, 187, 191–2, 194, 206, 208.
in Turkish 121, 132, 175, 208.
Haenisch, E. 181*, 192*.
Hamilton, J. R. 7*, 13, 29, 30, 88.
ḥarakāt 119, 120.
he 75.
Heissig, W. 186.
Henning, W. B. 11*, 29**, 76*, 101* ff., 106, 114.
Hephthalite alphabet 80.
HIiy. See *Hua-I i-yü.*
Houtsma, M. T. 192*.
Hsieh Yen-t'o people 28.
Hsien-pei language 27, 39 ff.
people 14, 20, 22 ff., 27, 39 ff.
Hsiung-nu people 6, 8ff.; and see Huns.
hu 6, 8.
Hua-I i-yü 126, 191, 217 ff.
Hu-han-hsieh *shan-yü* 22.
Huns language 40 ff.
people 11, 12, 14, 16 ff.; and see Hsiung-nu.

I Mongolian sound 190, 194, 205.
IB 68.
identification of letters 102.
Indo-European peoples 18, 21.
Iranian (Aramaic) alphabets 76 ff.
Irk Bitig XI, 68, 72, 81, 164.

J in Turkish 168
Jarring, G. 129.
jod 75.
Juan-juan language 27, 87.
people 14, 26, 42.

Kağan, title 26.
kao-ch'ê confederation 27, 28, 42.
Karabalğasun inscription 68.
Karaim language 49, 165.
Karakhanid dynasty 30 ff., 46, 178, 256.
Karasuk period 18.
Karlgren, B. 4*.
Karluk people 30, 46, 69.
Kāşğarī Maḥmūd al- 1, 44, 46, 50*, 51, 53, 94, 107, 119 ff., 131 ff., 136, 137, 140, 151, 158, 161 ff., 176, 220.
Katanov, N. F. VI*.
Kazan Tatar language. See Tatar.
Kazax language V, 50.
KB 50.
KÇ 80.
Khakas language 130, 170.
Khakasian inscriptions 69.

Khazar language 41
people 19, 31, 41.
Kıpçak languages 46 ff., 49, 52, 151, 220.
people 12, 31.
Kirakos of Ganja 195.
Kırğız language V, 40, 44, 50, 69, 164.
people 12, 21, 30, 40, 69.
Kitan people 217, 243, 246.
Ki-yü *shan yü* 21.
Koman language 48.
koph 76.
köşek 54.
Koutrigouroi tribe 38.
Kruger, J. R. VII*.
KT 68.
Kubilay Khan 189.
Kül Tégin 68, 89.
Küli Çor 68, 88.
Kun tribe 12.
Kutadğu Bilig I, 46, 48, 50, 66, 120, 178, 183.
Kyuner, N. V. 7*.
Kyzlasov, L. R. 70*, 72.

L- initial, in Turkish 172.
lamed 76, 104.
language, evolution of 33 ff., 211 ff.
Larichev, E. V. 17*.
Le Coq, A. von 80, 107*, 116**.
legal texts in Uyğur. See commercial *ditto*.
Ligdan Khan 186.
Ligeti, L. 124, 127*.
Liu Mau-tsai 7*, 8, 13, 28, 30, 88, 89.

M- initial, in Turkish languages 69, 91, 161, 171, 218.
McGovern, W. M. 7*.
Magyar language, Turkish loan words in 38, 52, 126 ff.
Malov, S. Ye. 68***.
1951 68.
1952 68 ff.
1959 68.
Manichaean Syriac alphabet 106, 114 ff.
Turkish texts 101, 106, 116 ff.
Manichaeism 113, 117, 118.
Mao-tun *shan yü* 4, 19 ff.
" mark " 177.
Marquart, J. 124*.
medical texts in Uyğur 81, 92, 101.
mem 76.
Mehmed II the Conqueror 185.
Ménage, V. L. IX.
Menander Protector 124.
Miller, R. A. 3*.
M.O.A. See Mongolian Official Alphabet.
Moghol language and people 214.
Mongols 10, 31, 43, 47, 52, 178, 185 ff., 245.
Mongolian language :—
basic vocabulary 10, 14, 16.
Chinese loan words in, see Chinese.
" Classical " 186, 196.
intervocalic hiatus in 196 ff., 206.
loan words in Turkish V, VII, 20, 52, 63, 184.
Persian loan words in 231 ff., 247.
structure of word 195 ff.
suffixes 201 ff.
suffixes, Turkish 201.
texts in Arabic transcription 192 ff.
Chinese transcription 191.
hP'ags-pa alphabet 189 ff.
M.O.A. 179 ff.
Turkish loan words in X, 37, 43, 44, 52, 125, 165, 193, 204, 217 ff.
unstable final consonants 198.
velar sounds 191, 193, 197 ff., 206.
Mongolian Official Alphabet 32, 48, 178 ff.
Moravcsik, Gy. 123*.
Muḥabbat Nāma 48.
Muḥammad Mahdī Xān XI*.
Muhannā, Camālu'l-dīn ibnu'l- 193*.
Mu-jung people 25 ff.
Müller, F. W. K. 96*, 108*, 114*, 175*.
Muqaddimatu'l-adab 193, 222 ff.

N- initial, in Turkish 171.
Nadzhip, E. N. 48*.
Nahcu'l-farādīs 20, 48, 163.
NK 97.
nun 76.
nuqṭa 176.

Ñ Turkish words containing 90–1, 161, 167.
ñ- initial, in early Turkish 126 ff., 160, 171, 228, 235.

-o-/-ö- in second syllable in early Turkish 95, 100, 164.
Oğuz languages 46, 49, 130, 151, 159 ff., 166–7, 170–1, 184, 218, 220.
peoples 12, 20, 28, 29, 31, 38, 40.
Oğuz Nāma 48, 52, 184.
okso 15, 235.
öküz 15, 208, 234.
Ölceytü, Ilkhan 182.
Ongin inscription X, 68, 70, 78, 82, 86.
Onogouroi tribe 38.
Opyt VII, 53, 66, 129.
ören 51.
Orkhon inscriptions 30, 68 ff., 118.
Orkun, H. N. 68*.
Osmanli language 49, 52–4, 120, 130, 140, 151, 155, 160 ff.
Ottuz Tatar. See Tatar.

Ougouroi tribe 28, 38, 123.
Outigouroi tribe 38.

P- initial, in Turkish 130, 160, 170.
hP'ags-pa alphabet VIII, X, 188 ff.
pai-tzu 183, 190.
Pavet de Courteille, A. 183**.
pe 76.
Peçeneg tribe 31.
P'ei Shih 3.
Pelliot, P. 87*.
PN 97.
polyphonic letters 55.
Poppe, N. N. 187*, 190*, 192*, 193*, 201*, 222*.
Poucha, P. 214*.
Priscus 123.
Pritsak, O. 125*.
Protobulgar. See Bulgar language.
Ptolemy geographer 124.
Pulleyblank, E. R. VIII, IX, 5, 6, 9, 10, 20, 28*, 29*, 88.

Quṭb 48*.

R- initial, in Turkish 172.
Rabğūzī 48*.
Radloff, V. V. III, VI*, VII*, 53, 66, 70*, 109*, 129*, 178, 183*.
Räsänen, M. VI*.
Raşīdu'l-dīn 47*.
Republican Turkish. See Osmanli.
resh 76.
hooked 107.
Rif'at, Kilisli Mu'allim 193*.
Rinchen, Y. 181*.
Ross, Sir E. Denison 102**.
Runic alphabet 67, 72 ff., 106, 162.
script, Turkish texts in 51, 53, 67 ff.

Saka documents and texts 86, 92, 123.
language and people 15, 16.
samech 76.
Sanglax V, IX, X, 48, 52–3, 192.
Saragouroi tribe 38.
Sarmatian people 124.
Saviroi tribe 20.
schin 76.
Secret History of the Mongols VII, 126, 191, 223 ff.
Selcuk dynasties 31, 46.
language 49.
Semitic alphabet, old 74, 114.
shan-yü 9, 19 ff.
Shcherbak, A. M. 48*, 68.
Simon, W. 5.
Sinkiang 16, 21 ff.
Sogdian alphabet 101 ff.
language 103 ff.
script, Turkish texts in 67, 101 ff.
Söz Derleme Dergisi VII, 129.
Staël-Holstein roll 10, 29, 92.
"standard" Turkish. See Turkish languages.
Stein, Sir Aurel 116.
Stönner, H. 92*.
Sucı inscription 68.
Sui Shu 8, 28.
Sulṭān Walad 49.
Suwarṇaprabhāsasūtra 96, 109.
Suwār tribe 20, 38.
SY 97.

Ş in Mongolian 191, 194, 206.
ş- initial, in Turkish 171.
Şine-usu inscription (*Şu*) 68.

"Taçam" 70.
Talas inscriptions 68, 72, 256.
T'ang Shu 28, 29.
Tan-shih-huai 24.
Tarduş confederation 28, 29, 89.
T'a-t'a T'ung-a 179.
Tatar languages and peoples 43–4, 46–7, 49, 220, 242.
tau 76.
Tavğaç language 37 ff., 125, 160, 217 ff.
people 13, 20, 25 ff., 39, 40.
Témür Kutluğ 185.
teŋri: 9, 10, 220, 223.
teth 75.
Thomas, F. W. VIII, X*, 5, 87*, 96, 122**.
Thomsen, V. 19, 72.
Tibetan script, Turkish texts *etc.* in 67, 86, 96 ff., 106, 122, 163.
transcriptions of Chinese X, 5, 123.
T'ieh-lê confederation 8, 20, 28, 42, 43, 109.
*Tiglig/Tigrig 21, 28, 40.
Ting-ling tribe 20, 28.
"Tokharian" language 15, 18, 92, 117, 227, 235.
Tokkuz Oğuz. See Oğuz.
Tölis confederation 12, 28, 29, 69.
Tomaschek 124.
ton 16.
Toñukuk inscription (*Toñ.*) 68.
t'o-pa. See Tavğaç people.
T'ou-man *shan-yü* 19.
Toussaint, Ch. 122*.
Toxtamış 185.
transcription alphabet for Turkish 57 ff.
"tribe" 11 ff.
Tsong-kha-pa 186.
TT. 91.
tümen 15, 227.
Tuna, O. N. 83*.
Tung Hu confederation 19 ff., 40.
Tungus languages and peoples 19, 36, 216.

Türgeş tribe 30, 69.
Türki language of Sinkiang 49, 208.
Turkish languages :—
Arabic loan words in V, 51, 176, 178.
basic vocabulary 14, 216.
Chinese loan words in. See Chinese.
dialectology 129.
double consonants in 81, 113, 116, 135, 136.
earliest surviving specimens of 1, 42, 45.
evolution of 40 ff.
Indian and Iranian loan words in 51, 77, 168, 172, 175.
lexicography of III, 53 ff., 62 ff.
loan words in Magyar. See Magyar language.
Mongolian. See Mongolian language.
l/r Turkish 37 ff., 40 ff., 44, 52, 124–7, 160, 164, 170–1, 218.
modern, evidential value of 128 ff.
Mongolian loan words in. See Mongolian language.
north eastern group V, 44, 50, 220.
north central group V, 50.
numerals in VI, X, 135, 150.
Persian loan words in V, 51, 178.
pre-eighth century V, 44, 135 ff., 140 ff., 160 ff.
Saka loan word in 16.
Sogdian loan words in. See Indian and Iranian *ditto*.
speaking peoples 1 ff.
"standard" Turkish 37 ff., 52, 126, 160–1, 220.
structure of word 135 ff.
suffixes in 45, 135, 140 ff.
Tokharian loan words in 15; and see **öküz, tümen.**
y- > c- languages 43, 44, 46, 127, 170, 220.
Türkmen language 48, 49, 52, 120, 130, 161 ff., 220.
peoples 31, 149.
Türkü language 42 ff., 46, 81, 118, 131, 143, 160 ff., 168, 218.
Manichaean, language 118, 161.
people 8, 9, 11, 14, 19, 27 ff., 41 ff., 68, 108, 118.
spelling of name 84 ff.
texts in 42, 68 ff.
Tuvan inscriptions 69, 163.
modern language V, VIII, 130.
tzaddi 76.

u. ("unvoiced") 60.
Ural River, name of 124.
Uyğur alphabet X, 67, 106 ff., 175 ff.
language VII, 43, 46–7, 52, 93, 95, 100–1, 131, 143, 151, 161, 168.
inscriptions in 68, 86.
texts in 46–7, 52–3, 67, 93, 100–1, 108, 117–8.
people 8, 29 ff., 101, 108.
Uyğur-A language and texts 43, 46, 53, 93, 100, 117–8, 161, 219.
Uzbek language 49, 54.
Uzun Ḥasan 185.

V 136.
v. ("voiced") 60.
v- initial in Turkish languages 170.
vau 75.
Vladimirtsov, B. Ya. 181*.
Volga Bulgar. See Bulgar language.
vowels Mongolian long 190, 193, 205.
Turkish final short 85 ff., 165.
long 82, 93, 119, 137 ff., 162 ff.
rounded in second syllable 95, 100, 164.
Vv 136.
VV 136.

W in Turkish 167.
Wei kingdom 25.
Northern, or Yüan 26, 41.
Wei Shu 3.
words, basic and elongated, in
Mongolian 195 ff.
Turkish 45, 135 ff.
Wu-huan people 20, 22 ff., 40.
Wu-sun people 18, 19.

X in Mongolian 194, 206.
Turkish 90, 167, 171, 173.
Xākānī language VII, 46–7, 50 ff., 94, 131, 143, 151, 161 ff., 168 ff., 184.
Xusrāw wa Şīrīn 48.
Xwarazm 48.
Xwarazmī 48.

Yağlakar clan 29.
Yakut language 130, 215.
yavğu: 28, 170.
yayık 124–5.
Yenisei inscriptions 68 ff.
Yoshitake, S. VIII, X*.
Yüeh-chih people 18, 19, 21.
Yule, Sir Henry 183*.
Yusuf Xaşş Ḥācib 50*.

Z lacking in Mongolian 191.
z- foreign initial, in Turkish 172.
zain 75
Zajączkowski, A. 48*.
Zamaxşarī, al- 193*.